GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 19, 2002, 21:31:06 ; Search time 2528.22 Seconds
(without alignments)
18247.064 Million cell updates/sec

Title: US-09-836-410A-2
Perfect score: 3418
Sequence: 1 caagtaacacccgcaagatg.....atgcaataaattgttttggg 3418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	950.2	27.8	2758	11	AK005056	AK005056 Mus muscu
	2	754.4	22.1	1093	10	BM456794	BM456794 AGENCOURT
	3	714.2	20.9	1959	11	AK007755	AK007755 Mus muscu
	4	704.8	20.6	821	10	BM044197	BM044197 603621442
c	5	679.2	19.9	704	10	BG067031	BG067031 H3049G08-
	6	673.4	19.7	689	9	BB616617	BB616617 BB616617
	7	656.4	19.2	659	9	BB659255	BB659255 BB659255
	8	649.8	19.0	675	9	AV270853	AV270853 AV270853
c	9	646.2	18.9	652	9	AW107262	AW107262 um13c03.x
	10	644.6	18.9	654	10	BF012472	BF012472 ux56g03.y
	11	629.6	18.4	673	9	BB577716	BB577716 BB577716
c	12	623.8	18.3	635	9	AW260482	AW260482 um80e10.x
	13	623.4	18.2	625	10	BG080108	BG080108 H3049G08-
c	14	617.8	18.1	710	9	AI744486	AI744486 wf89h01.x
	15	616.2	18.0	790	9	AU130763	AU130763 AU130763
	16	610	17.8	719	10	BF163411	BF163411 601771776
	17	608	17.8	617	9	BB478039	BB478039 BB478039
	18	607.4	17.8	744	10	BG671648	BG671648 DRNBWA09
	19	602.2	17.6	945	10	BF179047	BF179047 601806773
	20	601	17.6	978	10	BM463359	BM463359 AGENCOURT
	21	591	17.3	657	10	BF472586	BF472586 UI-M-BH3-
c	22	588.2	17.2	593	10	BG065941	BG065941 H3037B02-
	23	580	17.0	600	10	BG805362	BG805362 0691-69 M
	24	579	16.9	941	10	BG623888	BG623888 602649040
	25	575	16.8	603	10	BG079208	BG079208 H3037B02-
c	26	574.8	16.8	589	9	AW048763	AW048763 UI-M-BH1-
	27	573.4	16.8	626	10	BI662491	BI662491 603304484
	28	571.6	16.7	578	10	BI078250	BI078250 602872581
	29	561.4	16.4	629	10	BE300741	BE300741 ba90g11.x
c	30	559.2	16.4	622	9	AW554169	AW554169 L0236G10-
c	31	559	16.4	625	10	BM236189	BM236189 K0431E11-
c	32	545.8	16.0	549	10	BE852629	BE852629 uw26b05.x
c	33	545.2	16.0	684	9	AI177404	AI177404 EST221024
c	34	541.2	15.8	582	10	BM022195	BM022195 id60h03.x
c	35	539.6	15.8	615	10	BI966451	BI966451 id53e08.x
	36	537	15.7	549	10	BF453530	BF453530 maa69f12.
	37	535.2	15.7	931	10	BI687745	BI687745 603315658
	38	529.8	15.5	601	10	BF472327	BF472327 UI-M-BH3-
	39	528.8	15.5	666	10	D86662	D86662 D86662 Rat
	40	518.8	15.2	572	11	AK017653	AK017653 Mus muscu
	41	516.8	15.1	600	10	BI989826	BI989826 4051-81 M
	42	507.8	14.9	519	10	BF450556	BF450556 uz67h04.y
	43	506.8	14.8	510	9	AA435048	AA435048 ve06d04.r
	44	505.4	14.8	826	10	BI150077	BI150077 602849496
	45	505.2	14.8	511	10	BE626970	BE626970 uu08e01.y

ALIGNMENTS

RESULT 1
AK005056
LOCUS AK005056 2758 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300019C06:related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE, full insert sequence.
ACCESSION AK005056
VERSION AK005056.1 GI:12836717
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1300019C06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

```
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2758)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .2758
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="MGD:MGI:1895051"
                    /db_xref="taxon:10090"
                    /clone="1300019C06"
                    /sex="male"
                    /tissue_type="liver"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            64. .2658
                    /note="data source:SPTR, source key:Q9I8M2, evidence:ISS
                    putative
                    related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE"
                    /codon_start=1
                    /protein_id="BAB23782.1"
                    /db_xref="GI:12836718"
                    /translation="MPCVQLPAKESALFKRVLKCYEQKQYKNGLKFCKMILSNPKFAE
                    HGETLAMKGLILNCLGKREEAYEFVRKGLRSDVRSHVCWHVYGLLQRSDKKYDEAIKC
                    YRNALKLDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPTQRASWIGYAIAYHL
                    LKDYDTALKLLEEFRQTQQVPPNKIAYEYSELLLYQNQVMREANLFQESLEHIETYEK
                    LICDKLLVEEIKGEMLLKLGRLKEASEVFRNLIDWNAENWCYYEGLEKALQLRSLDER
                    LQLYEEVSKQHPRAVSPRRLPLSFAPGKKFRELMDKFLRPNFSKGCPPLFTTLKSLYC
                    DTEKVSIIQELVTNYEASLKMNGYFSPYENGEKEPPTTLIWVQYFLAQHYDKLGQYFL
                    ALEYINAVIASTPTLIELFYMKAKIYKHMGNLKEAAQWMDEAQSLDTADRFINSKCAK
                    YMLRANMIKEAEEMCSRFTREGTSAMENLNEMQCMWFETECISAYQRLGRYGDALKKC
                    HEVERHFLEITDDQFDFHTYCMRKMTLRAYVGLLRLEDALRRHTFYFKAARSAIEIYL
                    KLHDNPLTNDSKQQDIDSENLSAKEMKKMLSKQRRAQKKAKVEEERKHTERERQQKNQ
                    KKKREEEEEVTSGHKEELIPEKLERVDNPLEEAIKFLTPLKTLAAESIDTHLLAFEIY
                    FRKGKFLLMLQSVKRAFAIESNNPWLHECLIKFSKSVSNHSNLPDIVSKVLAQEMKKI
                    FVNKDLHSFNEDFLRHNATSLQHLLAGAKMMYFLDKSRQEKAIATATRLDETIKNKNV
                    KTLIKVSEALLDGSFGNCSSQYEEYRKTCHSLLPLTPAFLPAAREALGLSAELNHTAD
                    HKLLMNEI"
     polyA_signal   2741. .2746
                    /note="putative"
     polyA_site     2758
                    /note="putative"
BASE COUNT      902 a    528 c    615 g    713 t
ORIGIN
```

Query Match 27.8%; Score 950.2; DB 11; Length 2758;
Best Local Similarity 66.1%; Pred. No. 1.6e-152;
Matches 1372; Conservative 0; Mismatches 703; Indels 0; Gaps 0;

```
Qy      47 tggattggttatgctttaccatttattagaagactatgaaatggcagcaaaaattttaga 106
           ||||| | ||   ||| |||||  | |||| ||||| |  ||   |||  | |||||
Db     519 TGGATATGCCATCGCTTATCATTTGCTGAAAGATTATGATACAGCGCTAAAGCTATTAGA 578

Qy     107 agagtttaggaaaacacagcagacatctcctgataaagtggattatgaatatagtgaact 166
           ||| |||||  |||| |||||     |||| | ||| | | ||||||||||||||||||
Db     579 AGAATTTAGACAAACGCAGCAAGTTCCTCCCAACAAAATAGCTTATGAATATAGTGAACT 638

Qy     167 cctcttatatcagaatcaagttcttcgggaagcaggtctttatagagaagccctggaaca 226
             | ||||||||||||||||  | | || |||  | | | |   ||| |  |||||||
Db     639 GTTACTATATCAGAATCAAGTGATGAGAGAGGCAAATTTATTTCAGGAATCTTTGGAACA 698

Qy     227 tctttgtacctatgaaaagcagatttgtgataaacttgctgttgaagaaaccaaagggga 286
           | |    || ||||| |||  ||| ||||||||||||   || |||||||  || |||||
Db     699 TATAGAAACATATGAGAAGTTGATATGTGATAAACTTTTGGTGGAAGAAATTAAGGGGGA 758

Qy     287 acttctgttgcagttgtgtcgtttggaagatgctgctgacgtttatagaggattacaaga 346
           | | |||||| | ||| |  |  |  |||| ||  ||| || |  |||  ||   ||
Db     759 AATGCTGTTGAAATTGGGAAGACTAAAAGAAGCCAGTGAAGTATTCAGAAACTTGATTGA 818

Qy     347 gaggaatcctgaaaattgggcctattacaaaggcttagaaaaagcactgaagccagctaa 406
             ||||| ||||||||||   ||||| ||||||| ||||| ||| || | |    |
Db     819 TTGGAATGCTGAAAATTGGTGTTATTATGAAGGCTTGGAAAAGGCATTGCAACTTCGTTC 878

Qy     407 tatgttagaacggctaaaaatatatgaggaagcctggactaaataccccaggggactcgt 466
           | |    ||  |||| |  | |||||||||| | | |   |  |||||| | | |
Db     879 TCTAGACGAGAGGCTTCAGCTTTATGAGGAAGTCAGCAAGCAGCACCCCAGAGCAGTTTC 938

Qy     467 gccaagaaagctgcccttaaacttttatctggagagaagtttaaggagtgtttggatag 526
           || || | ||||||  | |  ||     | ||  |||||||||   ||     ||||||
Db     939 ACCCAGGAGGCTGCCGCTGAGTTTCGCCCCAGGTAAGAAGTTTCGAGAACTCATGGATAA 998

Qy     527 gttcctaaggatgaatttcagcaagggctgtccacctgtcttcaataccttgaggtcttt 586
           |||||| ||    || || ||||| || ||||||||| | |||| ||| |||| || ||
Db     999 GTTCCTGAGACCAAACTTTAGCAAAGGGTGTCCACCTCTGTTCACTACTTTGAAATCCTT 1058

Qy     587 atacagagataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttc 646
           ||  | |||| ||| |||||  ||||  |  | ||||| ||  |  ||||||| | ||
Db    1059 GTATTGTGATACAGAAAAGGTTTCAATAATCCAGGAACTTGTTACTAATTATGAAGCCTC 1118

Qy     647 tctaaaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccac 706
           ||| ||||    | ||   |||| ||| ||||  | ||  || ||||||| ||||| ||
Db    1119 TCTTAAAATGAATGGCTATTTTAGCCCTTATGAGAACGGGGAGAAGGAACCCCCAACAAC 1178

Qy     707 attactttgggtccagtactatttggcacagcattatgataaaattggtcagccatccat 766
             || | |||||||||| |  |||||||||||||||||||| |||| |||   |   |
Db    1179 TCTAATCTGGGTCCAGTATTTCCTGGCACAGCATTATGATAAACTTGGGCAGTATTTTCT 1238

Qy     767 tgctctggaatacataaatactgcaattgaaagtacaccaacattgatagaactcttttct 826
           ||| |||||||  || ||| ||| |||||  |||||  |||||  | ||||||||  ||
Db    1239 GGCTTTGGAATATATTAATGCTGTAATTGCTAGTACTCCAACTCTAATAGAACTATTCTA 1298

Qy     827 tgtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatgga 886
           | ||||| ||||| || ||||||   || ||| | |||||||| ||   |||||||||
Db    1299 CATGAAAGCAAAAATTTACAAGCATATGGGTAATCTCAAAGAAGCCGCACAGTGGATGGA 1358
```

```
Qy    887 tgaagcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatacat 946
          ||| || ||| |  ||||||| || ||||| || || |||||||| ||||| ||||||||
Db   1359 TGAGGCACAGTCTTTGGACACGGCTGACAGGTTCATCAATTCCAAATGTGCCAAATACAT 1418

Qy    947 gttaaaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaagg 1006
          | |   ||| ||  |||| ||||| || || |||||||| |||| ||| || |||||||||
Db   1419 GCTTCGAGCAAATATGATAAAAGAAGCAGAGGAAATGTGCTCCAGGTTCACAAGGGAAGG 1478

Qy   1007 aacttcagcggtagagaacctgaatgaaatgcagtgtatgtggtttcagacagagtgtgc 1066
          ||| || ||  | ||||| ||||||||||||||||||||||||||  |||| |||||
Db   1479 AACATCTGCCATGGAGAATCTGAATGAAATGCAGTGTATGTGGTTTGAGACGGAGTGCAT 1538

Qy   1067 tcaggcatacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattga 1126
          |  || ||  |     ||   | || ||| || ||  | || || || |||||| | ||
Db   1539 TTCTGCCTATCAGCGCCTGGGGAGATATGGGGATGCCTTGAAAAAGTGCCATGAAGTAGA 1598

Qy   1127 gagacattttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaa 1186
          ||| |||||| | || || |||||||| |||||||||| |||||||||| ||||| ||
Db   1599 GAGGCATTTTCTTGAGATAACCGATGATCAGTTTGACTTCCATACATACTGCATGAGAAA 1658

Qy   1187 gatcacccttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatcc 1246
          ||| |||||  |  | ||||| | | | || | | |||||||||  ||  ||   ||| |
Db   1659 GATGACCCTCCGTGCTTATGTTGGCCTCTTGAGATTAGAAGATGCTCTCAGAAGACATAC 1718

Qy   1247 attttacttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaaccc 1306
          ||||| ||||| || || |||   || ||||| || |||||||  | ||||| |||||
Db   1719 TTTTTATTTCAAGGCTGCTAGATCAGCAATTGAAATATATTTGAAATTACATGATAACCC 1778

Qy   1307 tctgacagatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagct 1366
          | | ||  |||| | |||| |||| ||    ||| ||| |||| |||| | ||||||  |
Db   1779 TTTAACCAATGACAGCAAACAACAAGACATAGATTCAGAAAACCTGTCAGCCAAAGAAAT 1838

Qy   1367 aaagaaactgcgtaataaacaaagaagagctcaaaagaaagcccagattgaagaagagaa 1426
          ||||||  ||| ||  || |||||||||||||| || || ||  || | ||||||||||
Db   1839 GAAGAAAATGCTTAGCAAGCAAAGAAGAGCTCAGAAAAAGGCTAAGGTAGAAGAAGAGAG 1898

Qy   1427 aaaaaatgccgaaaaagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatga 1486
          |||  |  | |||| |||    |  ||    || |  || || |||| ||  || || ||
Db   1899 AAAGCACACAGAAAGGGAACGCCAACAGAAAAACCAAAAGAAAAAAAGAGAAGAAGAAGA 1958

Qy   1487 cgaagaaattggaggccccaaagaagagcttatccctgagaaactggccaaggttgaaac 1546
          ||||  |   | || |  || ||||| ||||| ||||| |||||  |  | ||||||  |
Db   1959 AGAAGTTACCAGTGGTCATAAGGAAGAACTTATTCCTGAAAAACTAGAAAGGGTTGACAA 2018

Qy   1547 tccattggaagaagctattaagtttttaacaccattgaagaacttggtgaagaacaagat 1606
          |||||| |||||||| || |||||  | || ||  |||| | | | |      | |  ||
Db   2019 TCCATTAGAAGAAGCCATCAAGTTCCTCACTCCTCTGAAAACCCTTGCTGCTGAGAGCAT 2078

Qy   1607 agaaactcatctttttgcctttgagatctactttaggaaagaaaagtttcttttgatgct 1666
          || || || ||  | |||||||| || || ||||| ||||  ||||||  |  | ||| |
Db   2079 TGATACCCACCTGCTGGCCTTTGAAATATATTTTAGAAAAGGGAAGTTTTTGCTAATGTT 2138

Qy   1667 acaatcagtaaagcgggcatttgctattgattctagtcatccctggcttcatgagtgcat 1726
          || || || || || || ||||| |||||   || | |||| ||| | ||||| ||| |
Db   2139 GCAGTCTGTTAAACGAGCCTTTGCCATTGAAAGTAATAATCCTTGGTTACATGAATGCTT 2198

Qy   1727 gattcgactctttcattctgtgtgtgaaagtaaagacttacccgaaacagttagaacagt 1786
          ||||    | | | ||||| | | |    |   |  | || || |  || || | |||
Db   2199 GATTAAGTTTTCTAAATCTGTCTCTAATCACAGTAATCTTCCAGACATTGTGAGCAAAGT 2258

Qy   1787 attaaaacaagaaatgaatcgtctttttggagcaacaaatccaaagaattttaatgaaac 1846
            ||  ||||||||||  | ||||     | | ||     | |  ||||||||
Db   2259 TCTAGCTCAAGAAATGAAGAAGATATTTGTCAACAAAGATTTGCACAGCTTTAATGAGGA 2318

Qy   1847 ctttctgaaaaggaattctgattcattgccacatagattatcagctgccaaaatggtata 1906
          || || | |   ||| ||   ||  | |  ||| | |  ||| ||| |||||| | ||
Db   2319 TTTCCTCAGACACAATGCTACCTCTCTTCAGCATCTACTTGCAGGTGCTAAAATGATGTA 2378

Qy   1907 ttatttagattcttctagtcaaaaacgagcaatagagctggcgacaacacttgatggatc 1966
          || | | ||    || || ||| ||  |||||| |     || || |   | |||| | |
Db   2379 TTTTCTGGACAAGTCAAGGCAAGAAAAAGCAATTGCTACAGCCACTAGGTTGGATGAAAC 2438

Qy   1967 cctcaccaacagaaaccttcagacttgcatggaagtgttggaagccttgtgtgatggtag 2026
```

```
               | |  || | |||  |  |||||   ||  | ||||  || ||| |    ||||| ||
Db   2439 TATAAAAAATAAAAATGTGAAGACTCTAATAAAGGTGTCTGAGGCCCTCCTGGATGGCAG 2498

Qy   2027 cctacgagactgtaaagaagctgccgaagcctacagagcaagttgtcataagcttttccc 2086
          | |  |  ||||||           ||||  || ||    |  || ||||  ||  ||||
Db   2499 CTTTGGGAACTGTAGCTCCCAATATGAAGAGTATAGGAAGACCTGCCATAGCCTGCTCCC 2558

Qy   2087 ttatgctttggctttcatgcctcctggatacgaag 2121
               |    |||||| ||||| |||     ||||
Db   2559 ACTCACACCTGCTTTCCTGCCTGCTGCGAGGGAAG 2593


RESULT   2
BM456794
LOCUS       BM456794                1093 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6404070 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583707
            5', mRNA sequence.
ACCESSION   BM456794
VERSION     BM456794.1  GI:18505834
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1093)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12347  row: d  column: 12
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .1093
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583707"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      415 a    183 c    234 g    253 t      8 others
ORIGIN


  Query Match           22.1%;  Score 754.4;  DB 10;  Length 1093;
  Best Local Similarity   91.5%;  Pred. No. 5e-119;
  Matches  820;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    2;

Qy    683 tggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatta 742
          ||||||||||||||| ||||||||||||||||||||||||||||| |||||||| |||||
Db      1 TGGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATTA 60

Qy    743 tgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagtac 802
          ||| |||||||||||||||| |||||| |||| |||||||||||||| ||||||||||
Db     61 TGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTAC 120

Qy    803 accaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaatat 862
          ||| ||||| |||||||||||||| || ||||||||||||||||||||||| |||||
Db    121 ACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATAT 180

Qy    863 taaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagatttat 922
```

```
Db    181 TAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTAT 240

Qy    923 taattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaat 982

Db    241 CAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAAT 300

Qy    983 gtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagtg 1042

Db    301 GTGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGTG 360

Qy   1043 tatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaagc 1102

Db    361 CATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGC 420

Qy   1103 acttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttga 1162

Db    421 ACTTAAGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTGA 480

Qy   1163 ctttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaact 1222

Db    481 CTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTTATTAAAACT 540

Qy   1223 agaagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgagat 1282

Db    541 AGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGAT 600

Qy   1283 ctatttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgatac 1342

Db    601 CTATTTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGCTGATAC 660

Qy   1343 agcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaaa 1402

Db    661 AGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAA 720

Qy   1403 gaaagcccagattgaagaagagaaaaaaaatgccgaaaaagaaaagccgcaacggaatcc 1462

Db    721 GAAAGCCCAGATAGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCAGCAGAGAAATCA 780

Qy   1463 gaaaaagaaaaaggatgatgatgacgaagaaat-tggaggccccaaagaagagcttatcc 1521

Db    781 GAAAAAGAAGAAGGATGATGATGATGANGAGATAGGGAGGTCAAAAGAAAGAACTTATTC 840

Qy   1522 ctgagaaactggccaaggttgaaa--ctccattggaagaagctattaagtttttaa 1575

Db    841 CAGAGAAACTGGCCAAGGNTGAAAACTCCATTGGGAAGAAGCTATTAAATTTTTTA 896
```

RESULT 3
AK007755

LOCUS	AK007755 1959 bp mRNA linear HTC 19-JAN-2002
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810043G24:related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE, full insert sequence.
ACCESSION	AK007755
VERSION	AK007755.1 GI:12841504
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1810043G24.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (sites)
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2 (sites)
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (sites)
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites)
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 1959)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES	Location/Qualifiers
source	1. .1959 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGD:MGI:1895581" /db_xref="taxon:10090" /clone="1810043G24" /sex="male" /tissue_type="pancreas" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 day old"
misc_feature	1. .1959 /note="data source:SPTR, source key:Q9I8M2, evidence:ISS

related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE"
BASE COUNT 633 a 377 c 454 g 495 t
ORIGIN

Query Match 20.9%; Score 714.2; DB 11; Length 1959;
Best Local Similarity 69.0%; Pred. No. 3e-112;
Matches 977; Conservative 0; Mismatches 438; Indels 0; Gaps 0;

```
Qy      47 tggattggttatgctttaccatttattagaagactatgaaatggcagcaaaaattttaga 106
           |||||  |  ||   ||| ||||| |  |||| ||||| |  ||   |||  | |||||
Db     545 TGGATATGCCATCGCTTATCATTTGCTGAAAGATTATGATACAGCGCTAAAGCTATTAGA 604

Qy     107 agagtttaggaaaacacagcagacatctcctgataaagtggattatgaatatagtgaact 166
           ||| |||||  |||| || ||     |||| | ||| | | ||||||||||||||||||
Db     605 AGAATTTAGACAAACGCAACAAGTTCCTCCCAACAAAATAGCTTATGAATATAGTGAACT 664

Qy     167 cctcttatatcagaatcaagttcttcgggaagcaggtctttatagagaagccctggaaca 226
           | |||||||||||||||  | | || |||  | | | |   ||| |  |||||||
Db     665 GTTACTATATCAGAATCAAGTGATGAGAGAGGCAAATTTATTTCAGGAATCTTTGGAACA 724

Qy     227 tctttgtacctatgaaaagcagatttgtgataaacttgctgttgaagaaaccaaagggga 286
           | |    || ||||| |||  ||| ||||||||||||   || |||||||  || |||||
Db     725 TATAGAAACATATGAGAAGTTGATATGTGATAAACTTTTGGTGGAAGAAATTAAGGGGGA 784

Qy     287 acttctgttgcagttgtgtcgtttggaagatgctgctgacgtttatagaggattacaaga 346
           | | ||||||  | ||| |  |  |  |||| ||   ||| || |  |||   ||    ||
Db     785 AATGCTGTTGAAATTGGGAAGACTAAAAGAAGCCAGTGAAGTATTCAGAAACTTGATTGA 844

Qy     347 gaggaatcctgaaaattgggcctattacaaaggcttagaaaaagcactgaagccagctaa 406
           |||||  ||||||||||||   |||||  ||||||| ||||| ||| || | |    |
Db     845 TTGGAATGCTGAAAATTGGTGTTATTATGAAGGCTTGGAAAAGGCATTGCAACTTCGTTC 904

Qy     407 tatgttagaacggctaaaaatatatgaggaagcctggactaaataccccaggggactcgt 466
           | |    ||  |||| |  | ||||||||||| | | |   |  |||||||| | | |
Db     905 TCTAGACGAGAGGCTTCAGCTTTATGAGGAAGTCAGCAAGCAGCACCCCAGAGCAGTTTC 964

Qy     467 gccaagaaagctgcccttaaacttttatctggagagaagtttaaggagtgtttggatag 526
           || || | ||||||  | |  ||     | ||  ||||||||   ||    ||||||
Db     965 ACCCAGGAGGCTGCCGCTGAGTTTCGCCCCAGGTAAGAAGTTTCGAGAACTCATGGATAA 1024

Qy     527 gttcctaaggatgaatttcagcaagggctgtccacctgtcttcaataccttgaggtcttt 586
           ||||||  ||    || || ||||| || |||||||||  | ||||  ||| ||||  || ||
Db    1025 GTTCCTGAGACCAAACTTTAGCAAAGGGTGTCCACCTCTGTTCACTACTTTGAAATCCTT 1084

Qy     587 atacagagataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttc 646
           ||  | |||| ||| |||||  ||||  |  | |||||  ||   |  |||||||  | ||
Db    1085 GTATTGTGATACAGAAAAGGTTTCAATAATCCAGGAACTTGTTACTAATTATGAAGCCTC 1144

Qy     647 tctaaaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccac 706
           |||  ||||    | ||   |||| |||  ||||  | ||  || |||||||  ||||| ||
Db    1145 TCTTAAAATGAATGGCTATTTTAGCCCTTATGAGAACGGGGAGAAGGAACCCCCAACAAC 1204

Qy     707 attactttgggtccagtactatttggcacagcattatgataaaattggtcagccatccat 766
           ||  | ||||||||||| |   |||||||||||||||||||| |||| |||   |   |
Db    1205 TCTAATCTGGGTCCAGTATTTCCTGGCACAGCATTATGATAAACTTGGGCAGTATTTTCT 1264

Qy     767 tgctctggaatacataaatactgcaattgaaagtacaccaacattgatagaactctttct 826
           |||  |||||||  ||  |||  |||  |||||  |||||  |||||   | ||||||| ||
Db    1265 GGCTTTGGAATATATTAATGCTGTAATTGCTAGTACTCCAACTCTAATAGAACTATTCTA 1324

Qy     827 tgtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatgga 886
           |  |||||  |||||  ||  ||||||   ||  |||  | ||||||||  ||   |||||||||
Db    1325 CATGAAAGCAAAAATTTACAAGCATATGGGTAATCTCAAAGAAGCCGCACAGTGGATGGA 1384

Qy     887 tgaagcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatacat 946
           |||  ||  |||  |  |||||||  ||  |||||  ||  ||  ||||||||  |||||  ||||||||
Db    1385 TGAGGCACAGTCTTTGGACACGGCTGACAGGTTCATCAATTCCAAATGTGCCAAATACAT 1444

Qy     947 gttaaaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaagg 1006
           | |   |||  ||  ||||  |||||  ||  ||  ||||||||  ||||  |||  ||  ||||||||
Db    1445 GCTTCGAGCAAATATGATAAAAGAAGCAGAGGAAATGTGCTCCAGGTTCACAAGGGAAGG 1504
```

```
Qy    1007 aacttcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgc 1066
           |||  ||  ||  |  |||||  |||||||||||||||||||||||||||  ||||  |||||
Db    1505 AACATCTGCCATGGAGAATCTGAATGAAATGCAGTGTATGTGGTTTGAGACGGAGTGCAT 1564

Qy    1067 tcaggcatacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattga 1126
           |   ||  ||  |     ||    | ||  |||  ||  ||   | ||  ||  |||||  | ||
Db    1565 TTCTGCCTATCAGCGCCTGGGGAGATATGGGGATGCCTTGAAAAAGTGCCATGAAGTAGA 1624

Qy    1127 gagacattttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaa 1186
           |||  ||||||  |  ||  ||  ||||||||  ||||||||||  ||||||||||  |||||  ||
Db    1625 GAGGCATTTTCTTGAGATAACCGATGATCAGTTTGACTTCCATACATACTGCATGAGAAA 1684

Qy    1187 gatcacccttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatcc 1246
           |||  |||||  |  | |||||  |  |  |  ||  |  |  |||||||||  ||  ||   |||  |
Db    1685 GATGACCCTCCGTGCTTATGTTGGCCTCTTGAGATTAGAAGATGCTCTCAGAAGACATAC 1744

Qy    1247 attttacttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaaccc 1306
           |||||  |||||  ||  ||  |||   ||  |||||  ||  ||||||||   |  |||||  |||||
Db    1745 TTTTTATTTCAAGGCTGCTAGATCAGCAATTGAAATATATTTGAAATTACATGATAACCC 1804

Qy    1307 tctgacagatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagct 1366
           | | ||  ||||  | ||||  ||||  ||    |||  |||  ||||  ||||  |  ||||||   |
Db    1805 TTTAACCAATGACAGCAAACAACAAGACATAGATTCAGAAAACCTGTCAGCCAAAGAAAT 1864

Qy    1367 aaagaaactgcgtaataaacaaagaagagctcaaaagaaagcccagattgaagaagagaa 1426
           ||||||  |||  ||  ||  ||||||||||||||  ||  ||  ||  ||  |  ||||||||||
Db    1865 GAAGAAAATGCTTAGCAAGCAAAGAAGAGCTCAGAAAAAGGCTAAGGTAGAAGAAGAGAG 1924

Qy    1427 aaaaaatgccgaaaaagaaaagccgcaacggaatc 1461
           |||  |  | ||||  |||   |  ||   || |
Db    1925 AAAGCACACAGAAAGCGAACGCCAACAGAAAAACC 1959
```

```
RESULT   4
BM044197
LOCUS       BM044197                821 bp    mRNA    linear   EST 07-NOV-2001
DEFINITION  603621442F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5446979 5',
            mRNA sequence.
ACCESSION   BM044197
VERSION     BM044197.1  GI:16773464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1930  row: c  column: 12
            High quality sequence stop: 820.
FEATURES             Location/Qualifiers
     source          1. .821
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5446979"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
```

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 310 a 145 c 168 g 198 t
ORIGIN

Query Match 20.6%; Score 704.8; DB 10; Length 821;
Best Local Similarity 92.8%; Pred. No. 1.5e-110;
Matches 761; Conservative 0; Mismatches 57; Indels 2; Gaps 2;

```
Qy    594 gataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaa 653
          || ||||| |||||||||||| ||||||| ||||||| |||||||||| |||||||||
Db      2 GACAAAGAAAAGGTGGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAA 61

Qy    654 agttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactt 713
          || || ||  |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     62 AGCTGCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACATTACTT 121

Qy    714 tgggtccagtactatttggcacagcattatgataaaattggtcagccatccattgctctg 773
          |||||||||||||| |||||||| |||||||| |||||||||||||||||| |||||| ||
Db    122 TGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCAGCCATCTATTGCTTTG 181

Qy    774 gaatacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaa 833
          || |||||||||||||| |||||||||||||| ||||| |||||||||||||| || |||
Db    182 GAGTACATAAATACTGCTATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAA 241

Qy    834 gctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcc 893
          |||||||||||||||||||||| |||||||||||||||||| ||||||||||||| |||
Db    242 GCTAAAATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCC 301

Qy    894 caggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaa 953
          |||||| |||||||||||||||||||||| || ||||| |||||||||||||| |||||
Db    302 CAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAA 361

Qy    954 gccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttca 1013
          |||||||||||||||| |||||||||||||| || |||||||| |||||||||| |||
Db    362 GCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGAACATCA 421

Qy   1014 gcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggca 1073
          |||||||||||  |||||||||||||||| |||||||||| ||||| |||||
Db    422 GCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAACAGAATGTGCCCAGGCT 481

Qy   1074 tacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacat 1133
          || |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db    482 TATAAAGCAATGAATAAATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACAT 541

Qy   1134 tttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacc 1193
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||
Db    542 TTTATAGAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATTACC 601

Qy   1194 cttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatccatttttac 1253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTAC 661

Qy   1254 ttcaaagcagcgagaattgctattgagatcta-tttgaagcttcatgacaaccctctgac 1312
          ||||| ||||| |||||||||| |||||||| |||||||||||||||||| || ||
Db    662 TTCAAGGCAGCAAGAATTGCTATAGAGATCTATTTTGAAGCTTCATGACAACCCCCTTAC 721

Qy   1313 agatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaa 1372
          ||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db    722 AGATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAA 781

Qy   1373 actgcgtaataaacaaagaag-agctcaaaagaaagccca 1411
           || |||||||||||||||| ||||||||||||||||||
Db    782 GCTACGTAATAAACAAAGAAGAAGCTCAAAAGAAAGCCCA 821
```

RESULT 5
BG067031/c
LOCUS BG067031 704 bp mRNA linear EST 26-JAN-2001
DEFINITION H3049G08-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3049G08 3', mRNA sequence.
ACCESSION BG067031
VERSION BG067031.1 GI:12549600
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 704)
AUTHORS Kargul;G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka ,T.S., Carter,M.G. and Ko,M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3049G08-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3049 row: G column: 08
Seq primer: -21M13 Forward
High quality sequence stop: 704
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .704
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:H3049G08-3"
/db_xref="taxon:10090"
/clone="H3049G08"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
BASE COUNT 168 a 149 c 124 g 263 t
ORIGIN

Query Match 19.9%; Score 679.2; DB 10; Length 704;
Best Local Similarity 99.4%; Pred. No. 3.7e-106;
Matches 692; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy    729 ttggcacagcattatgataaaattggtcagccatccattgctctggaatacataaatact 788
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    695 TTTGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACATAAATACT 636

Qy    789 gcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaatctataag 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    635 GCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAG 576

Qy    849 catgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccctggacaca 908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     575 CATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACACA 516

Qy     909 gcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaa 968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     515 GCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGATTAAA 456

Qy     969 gaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtagagaacctg 1028
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 GAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAGAGAACCTG 396

Qy    1029 aatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaac 1088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 AATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAGCAATGAAC 336

Qy    1089 aaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatagaaatcacc 1148
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 AAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACC 276

Qy    1149 gatgaccagtttgactttcatacatactgtatgaggaagatcacccttagatcatatgtg 1208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 GATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGTG 216

Qy    1209 gacttattaaaactagaagatgtacttcgacagcatccattttacttcaaagcagcgaga 1268
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 GACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAGCAGCGAGG 156

Qy    1269 attgctattgagatctatttgaagcttcatgacaaccctctgacagatgagaacaaagaa 1328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 ATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGAACAAAGAA 96

Qy    1329 cacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaa 1388
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      95 CACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAA-TGCGTAATAAACAA 37

Qy    1389 agaagagctcaaaagaaagcccagattgaagaagag 1424
           |||||||||||||||||||||||| |||||||||
Db      36 AGAAGAGCTCAAAAGAAAGCCCAGATAGAAGAAGAG 1
```

```
RESULT   6
BB616617
LOCUS       BB616617                689 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB616617 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 4932426A01 5', mRNA sequence.
ACCESSION   BB616617
VERSION     BB616617.1  GI:16456621
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES                Location/Qualifiers
     source             1. .689
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="4932426A01"
                        /clone_lib="RIKEN full-length enriched, adult male testis"
                        /sex="male"
                        /tissue_type="testis"
                        /dev_stage="adult"
                        /lab_host="SOLR"
                        /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. Second strand cDNA was prepared with the
                        primer adapter of sequence [5'
                        GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
                        3']. cDNA was cloned into the XhoI and BamHI sites. "
BASE COUNT      223 a    126 c    143 g    191 t      6 others
ORIGIN
```

```
  Query Match           19.7%;  Score 673.4;  DB 9;  Length 689;
  Best Local Similarity   98.3%;  Pred. No. 3.6e-105;
  Matches  677;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    1510 aagagcttatccctgagaaactggccaaggttgaaactccattggaagaagctattaagt 1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAGCTATTAAGT 60

Qy    1570 ttttaacaccattgaagaacttggtgaagaacaagatagaaactcatctttttgcctttg 1629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTTTTGCCTTTG 120

Qy    1630 agatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagcgggcatttg 1689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGCGGGCATTTG 180

Qy    1690 ctattgattctagtcatccctggcttcatgagtgcatgattcgactctttcattctgtgt 1749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGT 240

Qy    1750 gtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaatgaatcgtc 1809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 GTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCGTC 300

Qy   1810 tttttggagcaacaaatccaaagaattttaatgaaacctttctgaaaaggaattctgatt 1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGAATTCTGATT 360

Qy   1870 cattgccacatagattatcagctgccaaaatggtatattatttagattcttctagtcaaa 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTTCTAGTCAAA 420

Qy   1930 aacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaaaccttcaga 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTCACCAACAGAAACCTTCAGA 480

Qy   1990 cttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctg 2049
          |||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||
Db    481 CTTGCATGGAGGTGTTGGAAGCCTTGTGTGATGGTAGCCTAGGAGACTGTAAAGAAGCTG 540

Qy   2050 ccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctttcatgcctc 2109
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    541 CCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTNTCATGCCTN 600

Qy   2110 ctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcagaaacggaag 2169
          ||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    601 CTGNATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGNTCTGCAGAAACGGAAG 660

Qy   2170 aactggccaatgaaatctgaacatcatta 2198
          |||||||       ||||||||||||||
Db    661 AACTGGCNNCATGAATCTGAACATCATTA 689
```

```
RESULT   7
BB659255
LOCUS       BB659255                659 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB659255 RIKEN full-length enriched, 13 days embryo heart Mus
            musculus cDNA clone D330028A07 5', mRNA sequence.
ACCESSION   BB659255
VERSION     BB659255.1  GI:16493078
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D330028A07"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     heart"
                     /tissue_type="heart"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      215 a    116 c    154 g    174 t
ORIGIN
```

```
  Query Match           19.2%;  Score 656.4;  DB 9;  Length 659;
  Best Local Similarity  99.8%;  Pred. No. 2.9e-102;
  Matches  657;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    279 aaaggggaacttctgttgcagttgtgtcgtttggaagatgctgctgacgtttatagagga 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 AAAGGGGAACTTCTGTTGCAGTTGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGA 61

Qy    339 ttacaagagaggaatcctgaaaattgggcctattacaaaggcttagaaaaagcactgaag 398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTACAAGAGAGGAATCCTGAAAATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAG 121

Qy    399 ccagctaatatgttagaacggctaaaaatatatgaggaagcctggactaaatacccccagg 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 CCAGCTAATATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGG 181

Qy    459 ggactcgtgccaagaaagctgcccttaaacttttatctggagagaagtttaaggagtgt 518
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    182 GGACTCGTGCCAAGAAGGCTGCCCTTAAACTTTTATCTGGAGAGAAGTTTAAGGAGTGT 241

Qy    519 ttggataggttcctaaggatgaatttcagcaagggctgtccacctgtcttcaataccttg 578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 TTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTG 301

Qy    579 aggtctttatacagagataaagagaaggtggcaatcgtagaagaactagtagttggttat 638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 AGGTCTTTATACAGAGATAAAGAGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTAT 361

Qy    639 gaaacttctctaaaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacct 698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     362 GAAACTTCTCTAAAAAGTTGTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCT 421

Qy     699 ccaaccacattactttgggtccagtactatttggcacagcattatgataaaattggtcag 758
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 CCAACCACATTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCAG 481

Qy     759 ccatccattgctctggaatacataaatactgcaattgaaagtacaccaacattgatagaa 818
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATAGAA 541

Qy     819 ctctttcttgtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccagg 878
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 CTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGG 601

Qy     879 tggatggatgaagcccaggccctggacacagcagacagatttattaattccaagtgtg 936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 TGGATGGATGAAGCCCAGGCCCTGGACACAGCAGACAGATTTATTAATTCCAAGTGTG 659
```

RESULT 8
AV270853

```
LOCUS       AV270853                675 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV270853 RIKEN full-length enriched, adult male testis (DH10B) Mus
            musculus cDNA clone 4931401K15 3', mRNA sequence.
ACCESSION   AV270853
VERSION     AV270853.2  GI:16389917
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 5, 1999 this sequence version replaced gi:6258890.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .675
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="4931401K15"
                      /clone_lib="RIKEN full-length enriched, adult male testis
                      (DH10B)"
                      /sex="male"
                      /tissue_type="testis"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. Second strand cDNA was prepared with the
                      primer adapter of sequence [5'
                      GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                      was cloned into the XhoI and BamHI sites. Vector: a
                      modified pBluescript KS(+) after bulk excision from Lambda
                      FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT      220 a    123 c    141 g    191 t
ORIGIN
```

```
  Query Match             19.0%;  Score 649.8;  DB 9;  Length 675;
  Best Local Similarity   98.7%;  Pred. No. 3.8e-101;
  Matches  666;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    1;

Qy    1553 ggaagaagctattaag--tttttaacaccattgaagaacttggtgaagaacaagatagaa 1610
           | |||||| |||||||  |||||| | ||||||||||||||||||||||||||||||||||
Db       1 GAAAGAAGTTATTAAGTTTTTTTACCCCCATTGAAGAACTTGGTGAAGAACAAGATAGAA 60

Qy    1611 actcatctttttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaa 1670
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAA 120

Qy    1671 tcagtaaagcgggcatttgctattgattctagtcatccctggcttcatgagtgcatgatt 1730
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATT 180

Qy    1731 cgactctttcattctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtatta 1790
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGACTCTTTCATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTA 240

Qy    1791 aaacaagaaatgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttt 1850
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAACAAGAAATGGATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTT 300

Qy    1851 ctgaaaaggaattctgattcattgccacatagattatcagctgccaaaatggtatattat 1910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTGAAAAGGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTAT 360

Qy    1911 ttagattcttctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctc 1970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTC 420

Qy    1971 accaacagaaaccttcagacttgcatggaagtgttggaagccttgtgtgatggtagccta 2030
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
```

```
Db    421 ACCAACAGAAACCTTCAGACTTGCATGGAGGTGTTGGAAGCCTTGTGTGATGGTAGCCTA 480

Qy   2031 cgagactgtaaagaagctgccgaagcctacagagcaagttgtcataagcttttccccttat 2090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGAGACTGTAAAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTAT 540

Qy   2091 gctttggctttcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagat 2150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTTTGGCTTTCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGAT 600

Qy   2151 agttctgcagaaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatg 2210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGTTCTGCAGAAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATG 660

Qy   2211 gaatgactttggacc 2225
          |||||||||||||||
Db    661 GAATGACTTTGGACC 675


RESULT  9
AW107262/c
LOCUS       AW107262                652 bp    mRNA    linear   EST 20-OCT-1999
DEFINITION  um13c03.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2192164 3' similar to WP:Y50D7_164.A CE22298 ;, mRNA
            sequence.
ACCESSION   AW107262
VERSION     AW107262.1  GI:6078062
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1004616
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2192164"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      155 a    138 c    114 g    244 t      1 others
ORIGIN


  Query Match           18.9%;  Score 646.2;  DB 9;  Length 652;
  Best Local Similarity   99.4%;  Pred. No. 1.6e-100;
  Matches  648;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    765 attgctctggaatacataaatactgcaattgaaagtacaccaacattgatagaactcttt 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    652 ATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTT 593

Qy    825 cttgtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatg 884
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    592 CTTGTAAAAGCTANAATCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATG 533

Qy    885 gatgaagcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatac 944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GATGAAGCCCAGGCCCTGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATAC 473

Qy    945 atgttaaaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaa 1004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    472 ATGTTAAAAGCCAACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAA 413

Qy   1005 ggaacttcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgt 1064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    412 GGAACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGT 353

Qy   1065 gctcaggcatacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaatt 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 GCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATT 293

Qy   1125 gagagacattttatagaaatcaccgatgaccagtttgactttcatacatactgtatgagg 1184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    292 GAGAGACATTTTATAGAAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGG 233

Qy   1185 aagatcacccttagatcatatgtggacttattaaaactagaagatgtacttcgacagcat 1244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    232 AAGATCACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCAT 173

Qy   1245 ccattttacttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaac 1304
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    172 CCATTTTACTTCAAAGCAGCGAGGATTGCTATTGAGATCTATTTGAAGCTTCATGACAAC 113

Qy   1305 cctctgacagatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagag 1364
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    112 CCTCTGACAAATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAG 53

Qy   1365 ctaaagaaactgcgtaataaacaaagaagagctcaaaagaaagcccagattg 1416
          |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     52 CTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATAG 1


RESULT  10
BF012472
LOCUS       BF012472                654 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  ux56g03.y1 Soares_NKWMD_mandible Mus musculus cDNA clone
            IMAGE:3514324 5' similar to TR:Q9VWI2 Q9VWI2 CG12202 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BF012472
VERSION     BF012472.1  GI:10712747
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1397204
            Seq primer: -40RP from Gibco
            High quality sequence stop: 471.
FEATURES              Location/Qualifiers
     source           1. .654
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3514324"
                      /clone_lib="Soares_NKWMD_mandible"
                      /tissue_type="mandible"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       228 a    118 c    137 g    168 t      3 others
ORIGIN


  Query Match           18.9%;  Score 644.6;  DB 10;  Length 654;
  Best Local Similarity  98.9%;  Pred. No. 3e-100;
  Matches  647;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     666 tttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtccagtac 725
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCCAGTAC 60

Qy     726 tatttggcacagcattatgataaaattggtcagccatccattgctctggaatacataaat 785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACATAAAT 120

Qy     786 actgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaatctat 845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTAT 180

Qy     846 aagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccctggac 905
           |||||||||||||||||||||||     ||||||||||||||||||||||||||||||||
Db     181 AAGCATGCTGGGAATATTAAAGAANGCTGCAGGTGGATGGATGAAGCCCAGGCCCTGGAC 240

Qy     906 acagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacctgatt 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGATT 300

Qy     966 aaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtagagaac 1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAGAGAAC 360

Qy    1026 ctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaagcaatg 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAGCAATG 420

Qy    1086 aacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatagaaatc 1145
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     421 AACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTNTATAGAAATC 480

Qy    1146 accgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagatcatat 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATAT 540

Qy    1206 gtggacttattaaaactagaagatgtacttcgacagcatccattttacttcaaagcagcg 1265
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     541 GTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTNTACTTCAAAGCAGCG 600

Qy    1266 agaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgag 1319
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAG 654
```

```
RESULT  11
BB577716
LOCUS       BB577716                673 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB577716 RIKEN full-length enriched, adult male medulla oblongata
            Mus musculus cDNA clone 6330400I15 5', mRNA sequence.
ACCESSION   BB577716
VERSION     BB577716.2  GI:16449448
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11474260.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .673
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="6330400I15"
                      /clone_lib="RIKEN full-length enriched, adult male medulla
                      oblongata"
                      /sex="male"
```

/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 213 a 118 c 130 g 212 t
ORIGIN

Query Match 18.4%; Score 629.6; DB 9; Length 673;
Best Local Similarity 98.4%; Pred. No. 1.1e-97;
Matches 657; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

```
Qy   2093 tttggctttcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatag 2152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 TTTGGCTTTCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAG 61

Qy   2153 ttctgcagaaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatgga 2212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTCTGCAGAAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGA 121

Qy   2213 atgactttggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctctt 2272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGACTTTGGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTT 181

Qy   2273 acttacacagaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatc 2332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ACTTACACAGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATC 241

Qy   2333 ctgctgaaaactatatataaaatatctaacattacaggatataggttcagtttcttaaaa 2392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CTGCTGAAAACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAA 301

Qy   2393 aattaaaagctgctaaaattgaggggtttaaaag-aaaaaaaaatccgtatcctattcct 2451
          |||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||
Db    302 AATTAAAAGCTGCTAAAATTGAGAGGTTTAAAAGAAAAAAAAAATCCGTATCCTATTCCT 361

Qy   2452 accttcccttcccatgtttttaactaatttatataaaatctggaggctataacagctaac 2511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 ACCTTCCCTTCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAAC 421

Qy   2512 atagcaggtgtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatct 2571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ATAGCAGGTGTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCT 481

Qy   2572 cagacagcataaataaatgctgttttagcactggattctttcactgagcacaaagagttg 2631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 CAGACAGCATAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTG 541

Qy   2632 ttggggctttagcatctgcctgattctgttacggggttggtgattgaccataggaagtat 2691
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    542 TTGGGGCTTTAACATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTTT 601

Qy   2692 gcaatgtgaatcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactg 2751
          |||||||||||||||||| ||||||| |||||||||||||||||| |||| | ||||
Db    602 GCAATGTGAATCACTGTGTTCAGAGCCCGTTACAACACATGCTTGACGGTGTA-ATACTG 660
```

```
Qy   2752 ggacacat 2759
          ||||||||
Db    661 GGACACAT 668
```

RESULT 12
AW260482/c
LOCUS AW260482 635 bp mRNA linear EST 23-DEC-1999
DEFINITION um80e10.x1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:2317674 3' similar to WP:Y50D7_164.A CE22298 ;, mRNA sequence.
ACCESSION AW260482
VERSION AW260482.1 GI:6633463
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 635)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: um80e10.y1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1010318
Seq primer: custom primer used
High quality sequence stop: 508.
FEATURES Location/Qualifiers
source 1. .635
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2317674"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG}, digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 152 a 137 c 111 g 235 t
ORIGIN

Query Match 18.3%; Score 623.8; DB 9; Length 635;
Best Local Similarity 98.9%; Pred. No. 1.1e-96;
Matches 628; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy    765 attgctctggaatacataaatactgcaattgaaagtacaccaacattgatagaactcttt 824
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    635 ATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTGT 576
```

```
Qy     825 cttgtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatg 884
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     575 CTTGTAAAAGCTACAATCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATG 516

Qy     885 gatgaagcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatac 944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     515 GATGAAGCCCAGGCCCTGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATAC 456

Qy     945 atgttaaaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaa 1004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 ATGTTAAAAGCCAACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAA 396

Qy    1005 ggaacttcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgt 1064
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 GGAACTTCTGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGT 336

Qy    1065 gctcaggcatacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaatt 1124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 GCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATT 276

Qy    1125 gagagacattttatagaaatcaccgatgaccagtttgactttcatacatactgtatgagg 1184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 GAGAGACATTTTATAGAAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGG 216

Qy    1185 aagatcacccttagatcatatgtggacttattaaaactagaagatgtacttcgacagcat 1244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 AAGATCACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCAT 156

Qy    1245 ccattttacttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaac 1304
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     155 CCATTTTACTTCAAAGCAGCGAGGATTGCTATTGAGATCTATTTGAAGCTTCATGACAAC 96

Qy    1305 cctctgacagatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagag 1364
           |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db      95 CCTCTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTAAGACAAAGAG 36

Qy    1365 ctaaagaaactgcgtaataaacaaagaagagctca 1399
           ||||||||||||||| ||||||||||||||||||
Db      35 CTAAAGAAACTGCGGAATAAACAAAGAAGAGCTCA 1

RESULT 13
BG080108
LOCUS       BG080108                625 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3049G08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3049G08 5', mRNA sequence.
ACCESSION   BG080108
VERSION     BG080108.1  GI:12562676
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3049G08-3
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3049  row: G  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 625
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3049G08-5"
                     /db_xref="taxon:10090"
                     /clone="H3049G08"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      211 a    111 c    143 g    160 t
ORIGIN


  Query Match           18.2%;  Score 623.4;  DB 10;  Length 625;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-96;
  Matches  624;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     422 aaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgcc 481
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1 AAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAGGCTGCC 60

Qy     482 cttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatgaa 541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAA 120

Qy     542 tttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaaga 601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGA 180

Qy     602 gaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtcg 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTCG 240

Qy     662 cctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcca 721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCCA 300

Qy     722 gtactatttggcacagcattatgataaaattggtcagccatccattgctctggaatacat 781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACAT 360

Qy     782 aaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaat 841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAAT 420

Qy     842 ctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccct 901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCT 480

Qy     902 ggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacct 961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCT 540

Qy     962 gattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtaga 1021
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAGA 600

Qy   1022 gaacctgaatgaaatgcagtgtatg 1046
          |||||||||||||||||||||||||
Db    601 GAACCTGAATGAAATGCAGTGTATG 625


RESULT 14
AI744486/c
LOCUS       AI744486                710 bp    mRNA    linear   EST 19-DEC-1999
DEFINITION  wf89h01.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2362801 3'
            similar to TR:O74985 O74985 N-TERMINAL ACETYLTRANSFERASE 1. ;, mRNA
            sequence.
ACCESSION   AI744486
VERSION     AI744486.1  GI:5112774
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 854   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2362801"
                     /clone_lib="NCI_CGAP_Co3"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization. "
BASE COUNT      170 a    146 c    121 g    267 t      6 others
ORIGIN


  Query Match             18.1%;  Score 617.8;  DB 9;  Length 710;
  Best Local Similarity   92.4%;  Pred. No. 1.1e-95;
  Matches  657;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy    714 tgggtccagtactatttggcacagcattatgataaaattggtcagccatccattgctctg 773
          |||||| ||||||| |||||||| |||||||| |||||||||||||||||| | |||| ||
Db    710 TGGGTCAAGTACTACTTGGCACACCATTATGACAAAATTGGTCAGCCATCTANTGCTNTG 651

Qy    774 gaatacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaa 833
          || |||||||||||||| ||||| |||||| | ||||| |||||  ||||||| || | |
Db    650 GAGTACATAAATACTGCTATTGANAGTACA-CTACATTAATAGACNTCTTTCTCGTGANA 592

Qy    834 gctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcc 893
          |||||||||||||||||||||||| |||||||||||||||| |||||||||||||| ||
Db    591 GCTAAAATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCN 532

Qy    894 caggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaa 953
          |||||| |||||||||||||||||||||| || ||||| ||||||||||||||| ||||||
Db    531 CAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAA 472

Qy    954 gccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttca 1013
          |||||||||||||||| |||||||||||||| || |||||||| |||||||||| |||
Db    471 GCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGAACATCA 412

Qy   1014 gcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggca 1073
          |||||||||||| |||||||||||||||| |||||||||| ||||| ||||| |||||
Db    411 GCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAACAGAATGTGCCCAGGCT 352

Qy   1074 tacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacat 1133
          || |||||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Db    351 TATAAAGCAATGAATAAATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACAT 292

Qy   1134 tttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacc 1193
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||| |||
Db    291 TTTATAGAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATTACC 232

Qy   1194 cttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatccattttac 1253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTAC 172

Qy   1254 ttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgaca 1313
          ||||| |||||| |||||||||| ||||||||||||||||||||||||||||| || |||
Db    171 TTCAAGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCCCTTACA 112

Qy   1314 gatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaa 1373
          |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    111 GATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAG 52

Qy   1374 ctgcgtaataaacaaagaagagctcaaaagaaagcccagattgaagaagag 1424
          || ||||||||||||||||||||||||||||||||||| |||||||||
Db     51 CTACGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATAGAAGAAGAG 1


RESULT 15
AU130763
LOCUS       AU130763                790 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  AU130763 NT2RP3 Homo sapiens cDNA clone NT2RP3001392 5', mRNA
            sequence.
ACCESSION   AU130763
VERSION     AU130763.1  GI:10991117
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone="NT2RP3001392"
                     /clone_lib="NT2RP3"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT      250 a    136 c    155 g    246 t      3 others
ORIGIN


  Query Match           18.0%;  Score 616.2;  DB 9;  Length 790;
  Best Local Similarity  90.0%;  Pred. No. 1.9e-95;
  Matches  718;  Conservative    0;  Mismatches   66;  Indels   14;  Gaps    5;

Qy   1882 gattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaatag 1941
           |||||||||||||||||||||||||||||| |||||| |||||||||| || ||||| ||||
Db      1 GATTATCAGCTGCCAAAATGGTATATTACTTAGATCCTTCTAGTCAGAAGCGAGCTATAG 60

Qy   1942 agctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatggaag 2001
           || |||| |||||||||||||| ||| ||||| |||||||||| ||||| || ||||| |
Db     61 AGTTGGCAACAACACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGGAGG 120

Qy   2002 tgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcctaca 2061
           | |||||||||||||| |||||||||||||| ||||||||||||||||||| |||   || |
Db    121 TATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAAATTTATA 180

Qy   2062 gagcaagttgtcataagcttttcccttatgctttggctttcatgcctcctggatacgaag 2121
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    181 GAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCCTCCTGGATATGAAG 240

Qy   2122 aggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggccaatg 2181
           |||||||||||||||||| || |||||||||||||||||| | ||||||||||||||||
Db    241 AGGATATGAAGATCACAGTTAATGGAGATAGTTCTGCAGAAGCTGAAGAACTGGCCAATG 300

Qy   2182 aaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtataat 2241
           |||| |||||||| ||||||||||||||||||||||||||||||||||||||| ||||||
Db    301 AAATTTGAACATCACTAAACAAGCAAATGGAATGACTTTGGACCATATCTAGTATATAAT 360

Qy   2242 atttttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaatgtt 2301
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    361 ATTTTTGTCACGCACCTGCTGCATTGCTCTAACTTACACAGAATGAGAGGAGTAAATGTT 420

Qy   2302 cttgccttcaaatagtc--ttacgttttttatcctgctg-aaaactatatataaaatatc 2358
           ||||||||||||||||   |||||||||||||||||| |||| ||||||||||||||||
Db    421 CTTGCCTTCAAATAGTGTTTTACGTTTTTTATCCTGCTGAAAAAGTATATATAAAATATC 480

Qy   2359 taacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgagggg 2418
           |||||||||||||| |||||||||||||||||||| |||||||||||||||||||| ||
Db    481 TAACATTACAGGATAGAGGTTCAGTTTCTTAAAAAATT-AAAGCTGCTAAAATTGAGTGG 539

Qy   2419 tttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaactaa 2478
           ||      |||||| |  || |||||||||||| ||  |||  ||    |||| ||||||
Db    540 TT------AAAAAAGATACCTTATCCTATTCCTCCCCACCCACCCATGTTTTTAAACTAA 593

Qy   2479 tttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatattac 2538
           |||||||||||||||||||| | |||||||||| ||||||||||||||||||||||||
Db    594 TTTATATAAAATCTGGAGGCTGTTACAGCTAACAAAGCAGGTGTGTGGCAGAAATATTAC 653

Qy   2539 tttaaatttgtcttgtgagattttgctatatctcagacagcataaataaatgctgtttta 2598
           |||||||||||||||||||||||| ||||||||||||||    |||||||||| ||||
Db    654 TTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC----ATAAATGCTGGTTTA 709

Qy   2599 gcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgattct 2658
           |||||||||||||||||||||||||||||||||| |||||||||||||||| |||  | |
Db    710 GCACTGGATTCTTTCACTGAGCACAAAGAGTTGNTGGGGCTTTAGCATCTGACTGGATTT 769

Qy   2659 gttacggggttggtgatt 2676
           | ||| ||| || |  ||
Db    770 GGTACNGGGNTGATTCTT 787


Search completed: July 20, 2002, 01:25:35

Job time: 14069 sec
```

OM of: US-09-836-410A-1 to: EST:* out_format : pfs

Date: Jul 20, 2002 3:07 AM

About: Results were produced by the GenCore software, version 4.5,.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_114913_29202/app_query.fasta_1.660
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09836410_@CGN1_1_2960
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1584.610000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_htc:AK005056	+	2008.50	3008.24	2.6e-158	2758	! AK005056 Mus musculus adult ma
gb_est2:BM456794	+	1483.00	2224.26	1.2e-114	1093	! BM456794 AGENCOURT_6404070 NIH
gb_est2:BM044197	+	1401.00	2103.21	6.7e-108	821	! BM044197 603621442F1 NIH_MGC_40
gb_htc:AK007755	+	1374.00	2054.67	3.4e-105	1959	! AK007755 Mus musculus 10 day c
gb_est2:BG067031	-	1180.00	1771.39	2.0e-89	704	! BG067031 H3049G08-3 NIA Mouse 1
gb_est1:AI744486	-	1176.00	1765.28	4.4e-89	710	! AI744486 wf89h01.x1 NCI_CGAP_Co
gb_est2:BG623888	+	1170.50	1754.45	1.8e-88	941	! BG623888 602649040F1 NIH_MGC_79
gb_est2:BF012472	+	1138.00	1708.73	6.3e-86	654	! BF012472 ux56g03.y1 Soares_NKWM
gb_est1:AW107262	-	1120.00	1681.61	2.0e-84	652	! AW107262 um13c03.x1 Sugano mous
gb_est1:BB616617	+	1111.00	1667.55	1.2e-83	689	! BB616617 BB616617 RIKEN full-le
gb_est2:BG080108	+	1093.00	1641.29	3.6e-82	625	! BG080108 H3049G08-5 NIA Mouse 1
gb_est1:AW260482	-	1081.00	1623.05	3.7e-81	635	! AW260482 um80e10.x1 Sugano mous
gb_est2:BE300741	+	1076.00	1615.60	9.6e-81	629	! BE300741 ba90g11.x2 NIH_MGC_5 H
gb_est1:AV227702	+	1072.00	1605.50	3.5e-80	988	! AV227702 AV227702 RIKEN full-le
gb_est2:BJ057588	+	1071.00	1607.93	2.6e-80	638	! BJ057588 BJ057588 NIBB Mochii r
gb_est1:BB478039	+	1065.00	1599.19	7.9e-80	617	! BB478039 BB478039 RIKEN full-le
gb_est1:AV270853	+	1051.00	1577.27	1.3e-78	675	! AV270853 AV270853 RIKEN full-le
gb_est2:BG805362	+	995.00	1493.90	5.8e-74	600	! BG805362 0691-69 Mouse E14.5 re
gb_est2:BI662491	+	993.00	1490.50	8.9e-74	626	! BI662491 603304484F1 NIH_CGAP_M
gb_est2:BE887347	+	980.50	1467.96	1.6e-72	943	! BE887347 601510175F1 NIH_MGC_71
gb_est2:BI445493	+	970.00	1455.60	7.9e-72	642	! BI445493 dae81d10.y3 NICHD XGC
gb_est1:AL594895	-	965.00	1448.89	1.9e-71	585	! AL594895 AL594895 XGC-gastrula
gb_est1:AW773112	+	962.00	1444.02	3.5e-71	608	! AW773112 EST00018 DT40subNB Gal
gb_est2:BI858089	+	960.50	1440.18	5.7e-71	725	! BI858089 603384214F1 NIH_MGC_87
gb_est2:BE867174	+	950.50	1426.00	3.5e-70	656	! BE867174 601442227F1 NIH_MGC_65
gb_est2:BM192037	+	948.00	1423.43	4.9e-70	574	! BM192037 dai23d09.y1 NICHD XGC
gb_est1:BB659255	+	927.00	1390.53	3.3e-68	659	! BB659255 BB659255 RIKEN full-le
gb_est2:BI687745	+	920.50	1377.62	1.7e-67	931	! BI687745 603315658F1 NCI_CGAP_M
gb_est2:BG079208	+	919.00	1379.27	1.4e-67	603	! BG079208 H3037B02-5 NIA Mouse 1
gb_est2:BE558139	+	909.00	1363.17	1.1e-66	675	! BE558139 fl20b04.y1 Zebrafish R
gb_est2:BM156001	-	904.00	1356.86	2.5e-66	589	! BM156001 fv15c04.x2 zebrafish a
gb_est2:BF938468	+	904.00	1356.23	2.7e-66	632	! BF938468 fm77a09.y1 Zebrafish a
gb_est1:AA435048	+	885.00	1329.51	8.3e-65	510	! AA435048 ve06d04.r1 Soares mous
gb_est2:BM183379	-	880.00	1320.77	2.5e-64	583	! BM183379 fv10h12.x2 zebrafish a
gb_est1:AJ392960	+	877.00	1315.31	5.1e-64	647	! AJ392960 AJ392960 dkfz426 Gallu
gb_est2:BJ042383	+	876.00	1315.91	4.7e-64	512	! BJ042383 BJ042383 NIBB Mochii r
gb_est2:BE626970	+	872.00	1309.89	1.0e-63	511	! BE626970 uu08e01.y2 Soares_mamm
gb_est2:BF450556	+	864.00	1297.69	4.9e-63	519	! BF450556 uz67h04.y1 NCI_CGAP_Ma
gb_est1:AA903456	-	859.00	1289.80	1.4e-62	540	! AA903456 ok58f08.s1 NCI_CGAP_Le
gb_est1:AW068237	-	857.00	1287.56	1.8e-62	495	! AW068237 cn23f10.y1 Normal Huma
gb_est2:BM005024	-	853.00	1280.11	4.7e-62	580	! BM005024 fu63b07.x1 zebrafis
gb_est2:BJ078440	-	846.50	1268.32	2.1e-61	723	! BJ078440 BJ078440 NIBB Mochi
gb_est2:BF047409	+	826.00	1238.99	9.1e-60	607	! BF047409 dc86h05.y1 NICHD XG
gb_est2:BM103338	-	809.00	1214.25	2.2e-58	550	! BM103338 fv20b04.x1 zebrafis
gb_est1:AA647535	+	804.00	1208.06	4.8e-58	473	! AA647535 vq80g07.s1 Knowles

seq_name: gb_htc:AK005056

seq_documentation_block:

LOCUS AK005056 2758 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300019C06:related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE, full insert sequence.
ACCESSION AK005056
VERSION AK005056.1 GI:12836717
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1300019C06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2758)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),

```
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES                Location/Qualifiers
     source             1. .2758
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="MGD:MGI:1895051"
                        /db_xref="taxon:10090"
                        /clone="1300019C06"
                        /sex="male"
                        /tissue_type="liver"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="adult"
     CDS                64. .2658
                        /note="data source:SPTR, source key:Q9I8M2, evidence:ISS
                        putative
                        related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE"
                        /codon_start=1
                        /protein_id="BAB23782.1"
                        /db_xref="GI:12836718"
                        /translation="MPCVQLPAKESALFKRVLKCYEQKQYKNGLKFCKMILSNPKFAE
                        HGETLAMKGLILNCLGKREEAYEFVRKGLRSDVRSHVCWHVYGLLQRSDKKYDEAIKC
                        YRNALKLDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPTQRASWIGYAIAYHL
                        LKDYDTALKLLEEFRQTQQVPPNKIAYEYSELLLYQNQVMREANLFQESLEHIETYEK
                        LICDKLLVEEIKGEMLLKLGRLKEASEVFRNLIDWNAENWCYYEGLEKALQLRSLDER
                        LQLYEEVSKQHPRAVSPRRLPLSFAPGKKFRELMDKFLRPNFSKGCPPLFTTLKSLYC
                        DTEKVSIIQELVTNYEASLKMNGYFSPYENGEKEPPTTLIWVQYFLAQHYDKLGQYFL
                        ALEYINAVIASTPTLIELFYMKAKIYKHMGNLKEAAQWMDEAQSLDTADRFINSKCAK
                        YMLRANMIKEAEEMCSRFTREGTSAMENLNEMQCMWFETECISAYQRLGRYGDALKKC
                        HEVERHFLEITDDQFDFHTYCMRKMTLRAYVGLLRLEDALRRHTFYFKAARSAIEIYL
                        KLHDNPLTNDSKQQDIDSENLSAKEMKKMLSKQRRAQKKAKVEEERKHTERERQQKNQ
                        KKKREEEEEVTSGHKEELIPEKLERVDNPLEEAIKFLTPLKTLAAESIDTHLLAFEIY
                        FRKGKFLLMLQSVKRAFAIESNNPWLHECLIKFSKSVSNHSNLPDIVSKVLAQEMKKI
                        FVNKDLHSFNEDFLRHNATSLQHLLAGAKMMYFLDKSRQEKAIATATRLDETIKNKNV
                        KTLIKVSEALLDGSFGNCSSQYEEYRKTCHSLLPLTPAFLPAAREALGLSAELNHTAD
                        HKLLMNEI"
     polyA_signal       2741. .2746
                        /note="putative"
     polyA_site         2758
                        /note="putative"
BASE COUNT      902 a    528 c    615 g    713 t
ORIGIN


alignment_scores:
            Quality: 2008.50              Length:    591
              Ratio:  3.915                Gaps:      1
 Percent Similarity:  86.802   Percent Identity: 62.775

alignment_block:
 US-09-836-410A-1 x AK005056   ..

 Align seg 1/1  to: AK005056  from: 1  to: 2758

      3 GluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLe 19
        |||||||||:::::|||||||||:::     :::::|||||:::::
    886 GAGAGGCTTCAGCTTTATGAGGAAGTCAGCAAGCAGCACCCCAGAGCAGT 935

     19 uValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysG 36
        :  ||||||:::|||||||||||:::|||      |||:::|||||:::|
    936 TTCACCCAGGAGGCTGCCGCTGAGTTTCGCCCCAGGTAAGAAGTTTCGAG 985

     36 luCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCysProPro 52
        ||   :::|||:::||||||||||   ||||||||||||||||||||||
    986 AACTCATGGATAAGTTCCTGAGACCAAACTTTAGCAAAGGGTGTCCACCT 1035

     53 ValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIl 69
        :::|||:::||||||:::|||||||||   |||   ||||||||||:::||
   1036 CTGTTCACTACTTTGAAATCCTTGTATTGTGATACAGAAAAGGTTTCAAT 1085

     69 eValGluGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgL 86
        |:::::|||||||||:::::|||||:::|||||||||
   1086 AATCCAGGAACTTGTTACTAATTATGAAGCCTCTCTTAAAATGAATGGCT 1135

     86 euPheAsnProAsnAspAspGlyLysGluGluProProThrThrLeuLeu 102
          |||:::|||   :::::::|||:::::::|||||||||||||||||:::
   1136 ATTTTAGCCCTTATGAGAACGGGGAGAAGGAACCCCCAACAACTCTAATC 1185

    103 TrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSe 119
        ||||||||||||:::|||||||||||||||||||||:::||||||
   1186 TGGGTCCAGTATTTCCTGGCACAGCATTATGATAAACTTGGGCAGTATTT 1235

    119 rIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuI 136
          :::|||||||||||||||||||:::|||   ||||||||||||||||
   1236 TCTGGCTTTGGAATATATTAATGCTGTAATTGCTAGTACTCCAACTCTAA 1285

    136 leGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIle 152
        ||||||||||   :::|||||||||||||||||||||   ||||||:::
   1286 TAGAACTATTCTACATGAAAGCAAAAATTTACAAGCATATGGGTAATCTC 1335

    153 LysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAs 169
        ||||||||||||:::|||||||||||||||||||||:::||||||||||||
   1336 AAAGAAGCCGCACAGTGGATGGATGAGGCACAGTCTTTGGACACGGCTGA 1385

    169 pArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuI 186
        ||||||||||||||||||||||||||||||||||:::||||||:::|
   1386 CAGGTTCATCAATTCCAAATGTGCCAAATACATGCTTCGAGCAAATATGA 1435

    186 leLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSer 202
        ||||||||||||||||||||||||||:::||||||||||||||||||||
   1436 TAAAAGAAGCAGAGGAAATGTGCTCCAGGTTCACAAGGGAAGGAACATCT 1485

    203 AlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCy 219
        |||:::|||||||||||||||||||||||||||||||||:::||||||||
   1486 GCCATGGAGAATCTGAATGAAATGCAGTGTATGTGGTTTGAGACGGAGTG 1535

    219 sAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysC 236
        |   :::|||||||:::   :::::::::::|||:::||||||||||||||
   1536 CATTTCTGCCTATCAGCGCCTGGGGAGATATGGGGATGCCTTGAAAAAGT 1585

    236 ysHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAsp 252
        ||||||||:::|||||||||||||:::|||||||||||||||||||||||
   1586 GCCATGAAGTAGAGAGGCATTTTCTTGAGATAACCGATGATCAGTTTGAC 1635

    253 PheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLe 269
        ||||||||||||||||||||||||||:::||||||||||:::|||||||   ||
   1636 TTCCATACATACTGCATGAGAAAGATGACCCTCCGTGCTTATGTTGGCCT 1685

    269 uLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaA 286
        ||||:::|||||||||:::|||||:::|||   ||||||||||||||||
   1686 CTTGAGATTAGAAGATGCTCTCAGAAGACATACTTTTTATTTCAAGGCTG 1735

    286 laArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThr 302
        |||||   ||||||||||||||||||||||||||||||||||||||||||
   1736 CTAGATCAGCAATTGAAATATATTTGAAATTACATGATAACCCTTTAACC 1785

    303 AspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGl 319
        :::::::::|||:::::::::   |||:::   |||:::|||   |||||
   1786 AATGACAGCAAACAACAAGACATAGATTCAGAAAACCTGTCAGCCAAAGA 1835
```

```
 319 uLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnI 336
1836 AATGAAGAAAATGCTTAGCAAGCAAAGAAGAGCTCAGAAAAAGGCTAAGG 1885

 336 leGluGluGluLysLysAsnAlaGluLysGluLysProGlnArgAsnPro 352
1886 TAGAAGAAGAGAGAAAGCACACAGAAAGGGAACGCCAACAGAAAAACCAA 1935

 353 LysLysLysLysAspAspAspAspGluGluIleGlyGlyProLysGluGl 369
1936 AAGAAAAAAAAGAGAAGAAGAAGAAGAAGTTACCAGTGGTCATAAGGAAGA 1985

 369 uLeuIleProGluLysLeuAlaLysValGluThrProLeuGluGluAlaI 386
1986 ACTTATTCCTGAAAAACTAGAAAGGGTTGACAATCCATTAGAAGAAGCCA 2035

 386 leLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluThr 402
2036 TCAAGTTCCTCACTCCTCTGAAAACCCTTGCTGCTGAGAGCATTGATACC 2085

 403 HisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeuMe 419
2086 CACCTGCTGGCCTTTGAAATATATTTTAGAAAAGGGAAGTTTTTGCTAAT 2135

 419 tLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisProTrpL 436
2136 GTTGCAGTCTGTTAAACGAGCCTTTGCCATTGAAAGTAATAATCCTTGGT 2185

 436 euHisGluCysMetIleArgLeuPheHisSerValCysGluSerLysAsp 452
2186 TACATGAATGCTTGATTAAGTTTTCTAAATCTGTCTCTAATCACAGTAAT 2235

 453 LeuProGluThrValArgThrValLeuLysGlnGluMetAsnArgLeuPh 469
2236 CTTCCAGACATTGTGAGCAAAGTTCTAGCTCAAGAAATGAAGAAGATATT 2285

 469 eGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysArgAsnS 486
2286 TGTCAACAAAGATTTGCACAGCTTTAATGAGGATTTCCTCAGACACAATG 2335

 486 erAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyrTyrLeu 502
2336 CTACCTCTCTTCAGCATCTACTTGCAGGTGCTAAAATGATGTATTTTCTG 2385

 503 AspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLeuAspGl 519
2386 GACAAGTCAAGGCAAGAAAAAGCAATTGCTACAGCCACTAGGTTGGATGA 2435

 519 ySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuGluAlaL 536
2436 AACTATAAAAAATAAAAATGTGAAGACTCTAATAAAGGTGTCTGAGGCCC 2485

 536 euCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAlaTyrArg 552
2486 TCCTGGATGGCAGCTTTGGGAACTGTAGCTCCCAATATGAAGAGTATAGG 2535

 553 AlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetProProGl 569
2536 AAGACCTGCCATAGCCTGCTCCCACTCACACCTGCTTTCCTGCCTGCTGC 2585

 569 yTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerAlaGluT 586
2586 G...AGGGAAGCCCTCGGTCTCAGTGCTGAGCTGAACCACACAGCTGACC 2632

 586 hrGluGluLeuAlaAsnGluIle 593
2633 ACAAGCTCTTAATGAATGAAATT 2655
```

seq_name: gb_est2:BM456794

seq_documentation_block:

LOCUS BM456794 1093 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6404070 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583707 5', mRNA sequence.
ACCESSION BM456794
VERSION BM456794.1 GI:18505834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1093)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12347 row: d column: 12
High quality sequence stop: 695.

FEATURES Location/Qualifiers
source 1. .1093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5583707"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 415 a 183 c 234 g 253 t 8 others
ORIGIN

alignment_scores:
Quality: 1483.00 Length: 336
Ratio: 4.678 Gaps: 8
Percent Similarity: 94.345 Percent Identity: 90.179

alignment_block:
US-09-836-410A-1 x BM456794 ..

Align seg 1/1 to: BM456794 from: 1 to: 1093

```
  93 GlyLysGluGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAl 109
   2 GGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGC 51

 109 aGlnHisTyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleA 126
  52 ACAACATTATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAA 101

 126 snThrAlaIleGluSerThrProThrLeuIleGluLeuPheLeuValLys 142
 102 ATACTGCTATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAA 151

 143 AlaLysIleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMe 159
 152 GCTAAAATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGAT 201

 159 tAspGluAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysC 176
 202 GGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAAT 251

 176 ysAlaLysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMet 192
 252 GTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATG 301
```

```
   193 CysSerLysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGl 209
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   302 TGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGA 351

   209 uMetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaM 226
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   352 AATGCAGTGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAA 401

   226 etAsnLysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHis 242
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   402 TGAATAAATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACAT 451

   243 PheIleGluIleThrAspAspGlnPheAspPheHisThrTyrCysMetAr 259
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   452 TTTATAGAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAG 501

   259 gLysIleThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspValL 276
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   502 GAAGATTACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTAC 551

   276 euArgGlnHisProPheTyrPheLysAlaAlaArgIleAlaIleGluIle 292
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   552 TTCGACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGATC 601

   293 TyrLeuLysLeuHisAspAsnProLeuThrAspGluAsnLysGluHisGl 309
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   602 TATTTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGA 651

   309 uAlaAspThrAlaAsnMetSerAspLysGluLeuLysLysLeuArgAsnL 326
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   652 AGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATA 701

   326 ysGlnArgArgAlaGlnLysLysAlaGlnIleGluGluGluLysLysAsn 342
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   702 AACAAAGAAGAGCTCAAAAGAAAGCCCAGATAGAAGAAGAGAAAAAAAAT 751

   343 AlaGluLysGluLysProGlnArgAsnProLysLysLysLysAspAspAs 359
       ||||||||||||||||  ||||||||  ||||||||||||||||||||||
   752 GCCGAAAAAGAAAAGCAGCAGAGAAATCAGAAAAAGAAGAAGGATGATGA 801

   359 pAspGluGluIle.GlyGlyProLysGluGluLeuIleProGluLysLeu 375
       ||||  ||||||  ||||||  |||:::|||||||||||||||||||||||
   802 TGATGANGAGATAGGGAGGTCAAAAGAAAGAACTTATTCCAGAGAAACTG 851

   376 AlaLysValGlu.ThrProLeu.GluGluAlaIleLys.PheLeuThrPr 391
       ||||||  |||  |||||||||  ||||||||||||||  ||||||||||
   852 GCCAAGGNTGAAAACTCCATTGGGAAGAAGCTATTAAATTTTTTAACACC 901

   391 oLeuLysAsn......LeuValLysAsnLysIleGluThr.HisLeuPhe 405
       |:::   :::      :::   :::::::::   :::||| |||||||||
   902 GGTGAANGAAACTTGGGTGAANGAACAAGAATAGAGGACCTCATCTTTTT 951

   406 Ala...PheGlu...IleTyrPheArgLysGluLysPheLeuLeuMet 419
       :::   ||||||   :::   :::||||||:::|||   :::|||::: 
   952 TGCCCCTTTGAGAATTTACCTTTAAGGAAAAGAAAAAGNTTTCTTCTT 999
```

seq_name: gb_est2:BM044197

seq_documentation_block:

```
LOCUS       BM044197                821 bp    mRNA    linear   EST 07-NOV-2001
DEFINITION  603621442F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5446979 5',
            mRNA sequence.
ACCESSION   BM044197
VERSION     BM044197.1  GI:16773464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1930  row: c  column: 12
            High quality sequence stop: 820.
FEATURES             Location/Qualifiers
     source          1. .821
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5446979"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      310 a    145 c    168 g    198 t
ORIGIN
```

alignment_scores:

```
            Quality: 1401.00              Length:    274
              Ratio:  5.151                 Gaps:      2
 Percent Similarity:  99.270   Percent Identity: 98.905
```

alignment_block:

US-09-836-410A-1 x BM044197 ..

Align seg 1/1 to: BM044197 from: 1 to: 821

```
    63 AspLysGluLysValAlaIleValGluGluLeuValValGlyTyrGluTh 79
       ||||||||||||||||||||:::||||||||||||||||||||||||||
     2 GACAAAGAAAAGGTGGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAAC 51

    79 rSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLysGluG 96
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    52 CTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATGATGGAAAGGAGG 101

    96 luProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyr 112
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   102 AACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATTAT 151

   113 AspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrAlaIl 129
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   152 GACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTAT 201

   129 eGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLysIleT 146
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   202 TGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCT 251

   146 yrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGluAla 162
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   252 ATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCC 301

   163 GlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaLysTy 179
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   302 CAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAAATA 351

   179 rMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerLysP 196
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   352 CATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGT 401
```

```
     196 heThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGlnCys 212
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGC 451

     213 MetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnLysPh 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     452 ATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATT 501

     229 eGlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIleGluI 246
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     502 TGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACATTTTATAGAAA 551

     246 leThrAspAspGlnPheAspPheHisThrTyrCysMetArgLysIleThr 262
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     552 TCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATTACC 601

     263 LeuArgSerTyrValAspLeuLeuLysLeuGluAspValLeuArgGlnHi 279
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     602 CTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCA 651

     279 sProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyr.LeuLys 295
         |||||||||||||||||||||||||||||||||||||||||| ||||||
     652 TCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGATCTATTTTGAAG 701

     296 LeuHisAspAsnProLeuThrAspGluAsnLysGluHisGluAlaAspTh 312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     702 CTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGCTGATAC 751

     312 rAlaAsnMetSerAspLysGluLeuLysLysLeuArgAsnLysGlnArgA 329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     752 AGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAA 801

     329 rg.AlaGlnLysLysAla 334
         || ||||||||||||||
     802 GAAGCTCAAAAGAAAGCC 819
```

seq_name: gb_htc:AK007755

seq_documentation_block:

LOCUS AK007755 1959 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810043G24:related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE, full insert sequence.

ACCESSION AK007755

VERSION AK007755.1 GI:12841504

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1810043G24.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 1959)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES Location/Qualifiers

source 1. .1959
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:1895581"
/db_xref="taxon:10090"
/clone="1810043G24"
/sex="male"
/tissue_type="pancreas"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 day old"

misc_feature 1. .1959
/note="data source:SPTR, source key:Q9I8M2, evidence:ISS related to PUTATIVE N-TERMINAL ACETYLTRANSFERASE"

BASE COUNT 633 a 377 c 454 g 495 t

ORIGIN

alignment_scores:

```
          Quality: 1374.00              Length:    349
            Ratio:  4.294                Gaps:      0
Percent Similarity:  91.691   Percent Identity: 70.201
```

```
alignment_block:
 US-09-836-410A-1 x AK007755   ..

 Align seg 1/1  to: AK007755  from: 1  to: 1959

       3 GluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLe 19
         |||||||||:::::|||||||||:::      :::::::||||:::::
     912 GAGAGGCTTCAGCTTTATGAGGAAGTCAGCAAGCAGCACCCCAGAGCAGT 961

      19 uValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysG 36
         :   ||||||:::|||||||||:::|||      |||:::|||||:::|
     962 TTCACCCAGGAGGCTGCCGCTGAGTTTCGCCCCAGGTAAGAAGTTTCGAG 1011

      36 luCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCysProPro 52
         ||   :::|||:::|||||||||   ||||||||||||||||||||||||
    1012 AACTCATGGATAAGTTCCTGAGACCAAACTTTAGCAAAGGGTGTCCACCT 1061

      53 ValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIl 69
         :::|||:::||||||:::|||||||||   |||   ||||||||||:::||
    1062 CTGTTCACTACTTTGAAATCCTTGTATTGTGATACAGAAAAGGTTTCAAT 1111

      69 eValGluGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgL 86
         |:::::|||||||||:::::|||||:::|||||||||
    1112 AATCCAGGAACTTGTTACTAATTATGAAGCCTCTCTTAAAATGAATGGCT 1161

      86 euPheAsnProAsnAspAspGlyLysGluGluProProThrThrLeuLeu 102
           |||:::|||   :::::::|||:::::::||||||||||||||||||:::
    1162 ATTTTAGCCCTTATGAGAACGGGGAGAAGGAACCCCCAACAACTCTAATC 1211

     103 TrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSe 119
         ||||||||||||:::|||||||||||||||||||||:::||||||
    1212 TGGGTCCAGTATTTCCTGGCACAGCATTATGATAAACTTGGGCAGTATTT 1261

     119 rIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuI 136
          :::|||||||||||||||||||:::::|||   |||||||||||||||
    1262 TCTGGCTTTGGAATATATTAATGCTGTAATTGCTAGTACTCCAACTCTAA 1311

     136 leGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIle 152
         ||||||||||   :::|||||||||||||||||||||   ||||||:::
    1312 TAGAACTATTCTACATGAAAGCAAAAATTTACAAGCATATGGGTAATCTC 1361

     153 LysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAs 169
         ||||||||||||:::|||||||||||||||||||:::|||||||||||||||
    1362 AAAGAAGCCGCACAGTGGATGGATGAGGCACAGTCTTTGGACACGGCTGA 1411

     169 pArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuI 186
         ||||||||||||||||||||||||||||||||||||||:::|||||||:::|
    1412 CAGGTTCATCAATTCCAAATGTGCCAAATACATGCTTCGAGCAAATATGA 1461

     186 leLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSer 202
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
    1462 TAAAAGAAGCAGAGGAAATGTGCTCCAGGTTCACAAGGGAAGGAACATCT 1511

     203 AlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCy 219
         |||:::|||||||||||||||||||||||||||||||||||:::||||||||
    1512 GCCATGGAGAATCTGAATGAAATGCAGTGTATGTGGTTTGAGACGGAGTG 1561

     219 sAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysC 236
         |   :::|||||||:::   ::::::::::|||:::|||||||||||||||
    1562 CATTTCTGCCTATCAGCGCCTGGGGAGATATGGGGATGCCTTGAAAAAGT 1611

     236 ysHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAsp 252
         ||||||||:::|||||||||||||:::|||||||||||||||||||||||||
    1612 GCCATGAAGTAGAGAGGCATTTTCTTGAGATAACCGATGATCAGTTTGAC 1661

     253 PheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLe 269
         ||||||||||||||||||||||||:::|||||||||:::||||||   ||
    1662 TTCCATACATACTGCATGAGAAAGATGACCCTCCGTGCTTATGTTGGCCT 1711

     269 uLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaA 286
         ||||:::|||||||||:::||||||:::|||   ||||||||||||||||
    1712 CTTGAGATTAGAAGATGCTCTCAGAAGACATACTTTTTATTTCAAGGCTG 1761

     286 laArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThr 302
         |||||   ||||||||||||||||||||||||||||||||||||||||||
    1762 CTAGATCAGCAATTGAAATATATTTGAAATTACATGATAACCCTTTAACC 1811

     303 AspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGl 319
         ::::::::::|||:::::::::   |||:::   |||:::|||   |||||
    1812 AATGACAGCAAACAACAAGACATAGATTCAGAAAACCTGTCAGCCAAAGA 1861

     319 uLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnI 336
         |:::|||||:::   :::||||||||||||||||||||||||||||||:::
    1862 AATGAAGAAAATGCTTAGCAAGCAAAGAAGAGCTCAGAAAAAGGCTAAGG 1911

     336 leGluGluGluLysLysAsnAlaGluLysGluLysProGlnArgAsn 351
         ::|||||||||:::|||:::::::|||:::|||:::   |||:::|||
    1912 TAGAAGAAGAGAGAAAGCACACAGAAAGCGAACGCCAACAGAAAAAC 1958

seq_name: gb_est2:BG067031

seq_documentation_block:
LOCUS       BG067031                704 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3049G08-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3049G08 3', mRNA sequence.
ACCESSION   BG067031
VERSION     BG067031.1  GI:12549600
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3049G08-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3049  row: G  column: 08
            Seq primer: -21M13 Forward
            High quality sequence stop: 704
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .704
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3049G08-3"
                     /db_xref="taxon:10090"
                     /clone="H3049G08"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
```

unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
Mol Genet 7: 1967-1978."
BASE COUNT 168 a 149 c 124 g 263 t
ORIGIN

alignment_scores:
Quality: 1180.00 Length: 231
Ratio: 5.130 Gaps: 0
Percent Similarity: 99.567 Percent Identity: 99.567

alignment_block:
US-09-836-410A-1 x BG067031/rev ..

Align seg 1/1 to reverse of: BG067031 from: 1 to: 704

```
109 AlaGlnHisTyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIl 125
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
692 GCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACAT 643

125 eAsnThrAlaIleGluSerThrProThrLeuIleGluLeuPheLeuValL 142
    ||||||||||||||||||||||||||||||||||||||||||||||||||
642 AAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAA 593

142 ysAlaLysIleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrp 158
    ||||||||||||||||||||||||||||||||||||||||||||||||||
592 AAGCTAAAATCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGG 543

159 MetAspGluAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLy 175
    ||||||||||||||||||||||||||||||||||||||||||||||||||
542 ATGGATGAAGCCCAGGCCCTGGACACAGCAGACAGATTTATTAATTCCAA 493

175 sCysAlaLysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluM 192
    ||||||||||||||||||||||||||||||||||||||||||||||||||
492 GTGTGCAAAATACATGTTAAAAGCCAACCTGATTAAAGAGGCTGAAGAAA 443

192 etCysSerLysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsn 208
    ||||||||||||||||||||||||||||||||||||||||||||||||||
442 TGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAGAGAACCTGAAT 393

209 GluMetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAl 225
    ||||||||||||||||||||||||||||||||||||||||||||||||||
392 GAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAGC 343

225 aMetAsnLysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgH 242
    ||||||||||||||||||||||||||||||||||||||||||||||||||
342 AATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGAC 293

242 isPheIleGluIleThrAspAspGlnPheAspPheHisThrTyrCysMet 258
    ||||||||||||||||||||||||||||||||||||||||||||||||||
292 ATTTTATAGAAATCACCGATGACCAGTTTGACTTTCATACATACTGTATG 243

259 ArgLysIleThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspVa 275
    ||||||||||||||||||||||||||||||||||||||||||||||||||
242 AGGAAGATCACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGT 193

275 lLeuArgGlnHisProPheTyrPheLysAlaAlaArgIleAlaIleGluI 292
    ||||||||||||||||||||||||||||||||||||||||||||||||||
192 ACTTCGACAGCATCCATTTTACTTCAAAGCAGCGAGGATTGCTATTGAGA 143

292 leTyrLeuLysLeuHisAspAsnProLeuThrAspGluAsnLysGluHis 308
    ||||||||||||||||||||||||||||||||||||||||||||||||||
142 TCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGAACAAAGAACAC 93

309 GluAlaAspThrAlaAsnMetSerAspLysGluLeuLysLysLeuArgAs 325
    ||||||||||||||||||||||||||||||||||||||||||   |||||
 92 GAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAATG.CGTAA 44

325 nLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGluGlu 339
    |||||||||||||||||||||||||||||||||||||||||||
 43 TAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATAGAAGAAGAG 1
```

seq_name: gb_est1:AI744486

seq_documentation_block:
LOCUS AI744486 710 bp mRNA linear EST 19-DEC-1999
DEFINITION wf89h01.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2362801 3' similar to TR:O74985 O74985 N-TERMINAL ACETYLTRANSFERASE 1. ;, mRNA sequence.
ACCESSION AI744486
VERSION AI744486.1 GI:5112774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 710)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 854 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 462.
FEATURES Location/Qualifiers
source 1. .710
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2362801"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "
BASE COUNT 170 a 146 c 121 g 267 t 6 others
ORIGIN

alignment_scores:
Quality: 1176.00 Length: 237
Ratio: 5.091 Gaps: 0
Percent Similarity: 97.468 Percent Identity: 96.203

alignment_block:
US-09-836-410A-1 x AI744486/rev ..

Align seg 1/1 to reverse of: AI744486 from: 1 to: 710

```
103 TrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSe 119
    ||||||:::|||||||||||||:::|||||||||||||||||||||||||
710 TGGGTCAAGTACTACTTGGCACACCATTATGACAAAATTGGTCAGCCATC 661

119 rIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuI 136
    |   |||  ||||||||||||||||||||||   ||||||  |||||||
660 TANTGCTNTGGAGTACATAAATACTGCTATTGANAGTACACT.ACATTAA 612

136 leGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIle 152
```

```
     ||:::   |||||||||   |||||||||||||||||||||||||||||||
  611 TAGACNTCTTTCTCGTGANAGCTAAAATCTATAAGCATGCTGGAAATATT 562

  153 LysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAs 169
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  561 AAAGAAGCTGCAAGGTGGATGGATGAGGCNCAGGCCTTGGACACAGCAGA 512

  169 pArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuI 186
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  511 CAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGA 462

  186 leLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSer 202
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  461 TTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGAACATCA 412

  203 AlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCy 219
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  411 GCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAACAGAATG 362

  219 sAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysC 236
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  361 TGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTAAGAAAT 312

  236 ysHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAsp 252
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  311 GTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTGAC 262

  253 PheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLe 269
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  261 TTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTT 212

  269 uLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaA 286
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  211 ATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAG 162

  286 laArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThr 302
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  161 CAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCCCTTACA 112

  303 AspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGl 319
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  111 GATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGACAAAGA 62

  319 uLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnI 336
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   61 GCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGA 12

  336 leGluGluGlu 339
      ||||||||||||
   11 TAGAAGAAGAG 1
```

seq_name: gb_est2:BG623888

seq_documentation_block:

```
LOCUS       BG623888                941 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602649040F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4770681 5',
            mRNA sequence.
ACCESSION   BG623888
VERSION     BG623888.1  GI:13675259
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1637  row: p  column: 10
            High quality sequence stop: 724.
FEATURES    Location/Qualifiers
     source 1. .941
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4770681"
            /clone_lib="NIH_MGC_79"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
            Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
            ); 5' and 3' adaptors were used in cloning as follows: 5'
            adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
            sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
            (where B = A, C, or G and N = A, C,  G, or T).  Average
            insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
            contained inserts  by PCR. This library was enriched for
            full-length clones and was constructed by Clontech
            Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
            Library."
BASE COUNT      358 a    172 c    211 g    200 t
ORIGIN
```

alignment_scores:

```
            Quality: 1170.50              Length:    286
              Ratio:  4.608                 Gaps:      5
 Percent Similarity:  88.811   Percent Identity: 82.168
```

alignment_block:

US-09-836-410A-1 x BG623888 ..

Align seg 1/1 to: BG623888 from: 1 to: 941

```
   95 GluGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHi 111
      :::|||||||||||||||||||||||||||||||||||||||||||||||
   45 AAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACA 94

  111 sTyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrA 128
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   95 TTATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTG 144

  128 laIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLys 144
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  145 CTATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAA 194

  145 IleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGl 161
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  195 ATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGA 244

  161 uAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaL 178
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  245 GGCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAA 294

  178 ysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSer 194
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  295 AATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCA 344

  195 LysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGl 211
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  345 AAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCA 394

  211 nCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnL 228
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  395 GTGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATA 444

  228 ysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIle 244
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  445 AATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACATTTTATA 494
```

```
     245 GluIleThrAspAspGlnPheAspPheHisThrTyrCysMetArgLysIl 261
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     495 GAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGAT 544

     261 eThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspValLeuArgG 278
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     545 TACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGAC 594

     278 lnHisProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyr.Le 294
         |||||||||||||||||||||||||||||||||||||||||||||||| ||
     595 AGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGATCTATTTT 644

     294 uLysLeuHisAspAsnProLeuThrAspGluAsnLysGluHisGlu.Ala 310
         ||||||||||||||:::|||||||||||||||||||||||||||||| |||
     645 GAAGCTTCATGACAAGCCCCTTACAGATGAGAATAAAGAACACGAAGGCT 694

     311 AspThrAlaAsnMetSerAspLysGluLeuLysLysLeuArgAsnLysGl 327
         ||||||   |||||||||:::::|||
     695 GATACAGAAAACATGTCTGAACAAGAG.....................TA 723

     327 nArgArgAla.........GlnLysLysAlaGlnIleGluGluGluLysL 341
          |||||||||          ::::::|||   :::::::::|||||||||:
     724 AAGAAGAGCTACGGTAATAAAACAAAAGAAAGAGGTTCAAGAAGAAAAGC 773

     341 ysAsnAlaGluLysGluLysProGlnArgAsnProLysLysLysLysAsp 357
         :::::      |||:::|||          |||   |||
     774 GCCAGGATTAGAAGCACAAACGAACAACAAACAGAAAAACACATATTGTG 823

     358 AspAspAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLy 374
               |||:::::::          |||:::::::|||   :::   :::::
     824 GGCCGCGACCAGAGG.........CCGCAGAAAGAGAGTGTTGAAAACCA 864

     374 sLeu 375
         :|||
     865 ACTC 868

seq_name: gb_est2:BF012472

seq_documentation_block:
LOCUS       BF012472                654 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  ux56g03.y1 Soares_NKWMD_mandible Mus musculus cDNA clone
            IMAGE:3514324 5' similar to TR:Q9VWI2 Q9VWI2 CG12202 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BF012472
VERSION     BF012472.1  GI:10712747
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1397204
            Seq primer: -40RP from Gibco
            High quality sequence stop: 471.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3514324"
                     /clone_lib="Soares_NKWMD_mandible"
                     /tissue_type="mandible"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      228 a    118 c    137 g    168 t      3 others
ORIGIN


alignment_scores:
            Quality: 1138.00              Length:    218
              Ratio:  5.269                Gaps:      0
 Percent Similarity:  99.083   Percent Identity: 98.165

alignment_block:
 US-09-836-410A-1 x BF012472   ...

 Align seg 1/1  to: BF012472  from: 1  to: 654

      87 PheAsnProAsnAspAspGlyLysGluGluProProThrThrLeuLeuTr 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 TTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTG 50

     103 pValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSerI 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCA 100

     120 leAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuIle 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATA 150

     137 GluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIleLy 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGGAATATTAA 200

     153 sGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAspA 170
         ||||:::::::||||||||||||||||||||||||||||||||||||||||
     201 AGAANGCTGCAGGTGGATGGATGAAGCCCAGGCCCTGGACACAGCAGACA 250

     170 rgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIle 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGATT 300

     187 LysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSerAl 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 AAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGC 350

     203 aValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCysA 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTG 400

     220 laGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysCys 236
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 CTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATGT 450

     237 HisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAspPh 253
         ||||||||||||||||||   ||||||||||||||||||||||||||||||
     451 CATGAAATTGAGAGACATTNTATAGAAATCACCGATGACCAGTTTGACTT 500

     253 eHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLeuL 270
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 TCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGTGGACTTAT 550

     270 euLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaAla 286
         ||||||||||||||||||||||||||||||||   ||||||||||||||||
     551 TAAAACTAGAAGATGTACTTCGACAGCATCCATTNTACTTCAAAGCAGCG 600

     287 ArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThrAs 303
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 AGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGA 650
```

```
     303 pGlu 304
         ||||
     651 TGAG 654

seq_name: gb_est1:AW107262

seq_documentation_block:
LOCUS       AW107262                652 bp    mRNA    linear   EST 20-OCT-1999
DEFINITION  um13c03.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2192164 3' similar to WP:Y50D7_164.A CE22298 ;, mRNA
            sequence.
ACCESSION   AW107262
VERSION     AW107262.1  GI:6078062
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1004616
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2192164"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      155 a    138 c    114 g    244 t      1 others
ORIGIN


alignment_scores:
            Quality: 1120.00             Length:    217
              Ratio:  5.185                Gaps:      0
 Percent Similarity:  99.539   Percent Identity: 99.078

alignment_block:
 US-09-836-410A-1 x AW107262/rev   ..

 Align seg 1/1  to reverse of: AW107262  from: 1  to: 652

    120 IleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuIl 136
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    652 ATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGAT 603

    136 eGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIleL 153
        |||||||||||||||||||||   ||||||||||||||||||||||||||
    602 AGAACTCTTTCTTGTAAAAGCTANAATCTATAAGCATGCTGGGAATATTA 553

    153 ysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAsp 169
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    552 AAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACACAGCAGAC 503

    170 ArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIl 186
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    502 AGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGAT 453

    186 eLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSerA 203
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    452 TAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAG 403

    203 laValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCys 219
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 CGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGT 353

    220 AlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysCy 236
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 GCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATG 303

    236 sHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAspP 253
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 TCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAGTTTGACT 253

    253 heHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLeu 269
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 TTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGTGGACTTA 203

    270 LeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaAl 286
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAGCAGC 153

    286 aArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThrA 303
        |||||||||||||||||||||||||||||||||||||||||||||||||:
    152 GAGGATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAA 103

    303 spGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGlu 319
        ::|||||||||||||||||||||||||||||||||||||||||||||||||
    102 ATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAG 53

    320 LeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIl 336
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 CTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGAT 3

    336 e 336
        |
      2 A 2

seq_name: gb_est1:BB616617

seq_documentation_block:
LOCUS       BB616617                689 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB616617 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 4932426A01 5', mRNA sequence.
ACCESSION   BB616617
VERSION     BB616617.1  GI:16456621
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
```

,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .689
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4932426A01"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. "

BASE COUNT 223 a 126 c 143 g 191 t 6 others
ORIGIN

alignment_scores:
Quality: 1111.00 Length: 222
Ratio: 5.096 Gaps: 0

Percent Similarity: .98.198 Percent Identity: 97.748

alignment_block:
US-09-836-410A-1 x BB616617 ..

Align seg 1/1 to: BB616617 from: 1 to: 689

```
    369 GluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluGluAl 385
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      3 GAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAGC 52

    385 aIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluT 402
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     53 TATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAA 102

    402 hrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeu 418
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    103 CTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTG 152

    419 MetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisProTr 435
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    153 ATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATCCCTG 202

    435 pLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSerLysA 452
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    203 GCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGTAAAG 252

    452 spLeuProGluThrValArgThrValLeuLysGlnGluMetAsnArgLeu 468
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    253 ACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCGTCTT 302

    469 PheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysArgAs 485
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    303 TTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGAA 352

    485 nSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyrTyrL 502
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    353 TTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATT 402

    502 euAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLeuAsp 518
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    403 TAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGAT 452

    519 GlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuGluAl 535
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    453 GGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAGGTGTTGGAAGC 502

    535 aLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAlaTyrA 552
        |||||||||||||||||||   |||||||||||||||||||||||||||
    503 CTTGTGTGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCCGAAGCCTACA 552

    552 rgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetProPro 568
        |||||||||||||||||||||||||||||||||||||   ||||||
    553 GAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTNTCATGCCTNCT 602

    569 GlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerAlaGl 585
        ||||||||||||||||||||||||||||||||||||||:::||||||||
    603 GNATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGNTCTGCAGA 652

    585 uThrGluGluLeuAla 590
        |||||||||||||||
    653 AACGGAAGAACTGGCN 668
```

seq_name: gb_est2:BG080108

seq_documentation_block:
LOCUS BG080108 625 bp mRNA linear EST 26-JAN-2001
DEFINITION H3049G08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3049G08 5', mRNA sequence.
ACCESSION BG080108
VERSION BG080108.1 GI:12562676
KEYWORDS EST.
SOURCE house mouse.

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3049G08-3
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3049  row: G  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 625
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3049G08-5"
                     /db_xref="taxon:10090"
                     /clone="H3049G08"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT       211 a    111 c    143 g    160 t
ORIGIN


alignment_scores:
            Quality: 1093.00              Length:    208
              Ratio:  5.255                 Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 99.519

alignment_block:
 US-09-836-410A-1 x BG080108   ..

 Align seg 1/1  to: BG080108  from: 1  to: 625

       6 LysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValProAr 22
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAG 51

      22 gLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysGluCysLeuA 39
         |:::||||||||||||||||||||||||||||||||||||||||||||||
      52 AAGGCTGCCCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGG 101

      39 spArgPheLeuArgMetAsnPheSerLysGlyCysProProValPheAsn 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 ATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGTCCACCTGTCTTCAAT 151

      56 ThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIleValGluGl 72
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGTGGCAATCGTAGAAGA 201

      72 uLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPheAsnP 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     202 ACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTCGCCTATTTAACC 251

      89 roAsnAspAspGlyLysGluGluProProThrThrLeuLeuTrpValGln 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCCAG 301

     106 TyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSerIleAlaLe 122
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     302 TACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCT 351

     122 uGluTyrIleAsnThrAlaIleGluSerThrProThrLeuIleGluLeuP 139
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCT 401

     139 heLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIleLysGluAla 155
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGGAATATTAAAGAAGCT 451

     156 AlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAspArgPheIl 172
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GCCAGGTGGATGGATGAAGCCCAGGCCCTGGACACAGCAGACAGATTTAT 501

     172 eAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIleLysGluA 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     502 TAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGATTAAAGAGG 551

     189 laGluGluMetCysSerLysPheThrArgGluGlyThrSerAlaValGlu 205
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     552 CTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAGAG 601

     206 AsnLeuAsnGluMetGlnCysMet 213
         |||||||||||||||||||||||||
     602 AACCTGAATGAAATGCAGTGTATG 625

seq_name:: gb_est1:AW260482

seq_documentation_block:
LOCUS       AW260482                635 bp    mRNA    linear   EST 23-DEC-1999
DEFINITION  um80e10.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:2317674 3' similar to WP:Y50D7_164.A CE22298 ;, mRNA
            sequence.
ACCESSION   AW260482
VERSION     AW260482.1  GI:6633463
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: um80e10.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
```

```
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          MGI:1010318
          Seq primer: custom primer used
          High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2317674"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      152 a    137 c    111 g    235 t
ORIGIN


alignment_scores:
           Quality: 1081.00              Length:    211
             Ratio:  5.197                 Gaps:      0
 Percent Similarity:  98.578   Percent Identity: 98.578

alignment_block:
 US-09-836-410A-1 x AW260482/rev   ..

 Align seg 1/1  to reverse of: AW260482  from: 1  to: 635

     120 IleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuIl 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     635 ATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGAT 586

     136 eGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIleL 153
         |||||||   ||||||||||||   |||||||||||||||||||||||||
     585 AGAACTCTGTCTTGTAAAAGCTACAATCTATAAGCATGCTGGGAATATTA 536

     153 ysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAsp 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 AAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACACAGCAGAC 486

     170 ArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIl 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     485 AGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACCTGAT 436

     186 eLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSerA 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     435 TAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCTG 386

     203 laValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCys 219
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     385 CGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGT 336

     220 AlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysCy 236
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     335 GCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATG 286

     236 sHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAspP 253
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     285 TCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAGTTTGACT 236

     253 heHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLeu 269
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     235 TTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGTGGACTTA 186

     270 LeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaAl 286
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 TTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAGCAGC 136

     286 aArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThrA 303
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     135 GAGGATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAG 86

     303 spGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGlu 319
         |||||||||||||||||||||||||||||||||||   |||||||||
      85 ATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTAAGACAAAGAG 36

     320 LeuLysLysLeuArgAsnLysGlnArgArgAla 330
         ||||||||||||||||||||||||||||||||
      35 CTAAAGAAACTGCGGAATAAACAAAGAAGAGCT 3

seq_name: gb_est2:BE300741

seq_documentation_block:
LOCUS       BE300741                 629 bp    mRNA    linear   EST 14-JUL-2000
DEFINITION  ba90g11.x2 NIH_MGC_5 Homo sapiens cDNA clone IMAGE:2924612 3'
            similar to WP:Y50D7_164.A CE22298 ;, mRNA sequence.
ACCESSION   BE300741
VERSION     BE300741.1  GI:9184428
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2924612"
                     /clone_lib="NIH_MGC_5"
                     /tissue_type="carcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: cervix; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      254 a    103 c    133 g    137 t      2 others
ORIGIN


alignment_scores:
           Quality: 1076.00              Length:    208
```

```
            Ratio:  5.249                Gaps:      0
Percent Similarity:  98.558   Percent Identity: 98.558

alignment_block:
US-09-836-410A-1 x BE300741   ..

Align seg 1/1  to: BE300741  from: 1  to: 629

    144 LysIleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAs 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      3 AAAATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGA 52

    160 pGluAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysA 177
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     53 TGAGGCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTG 102

    177 laLysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCys 193
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    103 CAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGC 152

    194 SerLysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMe 210
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    153 TCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAAT 202

    210 tGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetA 227
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    203 GCAGTGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGA 252

    227 snLysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHisPhe 243
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    253 ATAAATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACATTTT 302

    244 IleGluIleThrAspAspGlnPheAspPheHisThrTyrCysMetArgLy 260
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    303 ATAGAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAA 352

    260 sIleThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspValLeuA 277
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    353 GATTACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTC 402

    277 rgGlnHisProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyr 293
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    403 GACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGATCTAT 452

    294 LeuLysLeuHisAspAsnProLeuThrAspGluAsnLysGluHisGluAl 310
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    453 TTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGC 502

    310 aAspThrAlaAsnMetSerAspLysGluLeuLysLysLeuArgAsnLysG 327
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    503 TGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAAC 552

    327 lnArgArgAlaGlnLysLysAlaGlnIleGluGluGluLysLysAsnAla 343
        ||||||||||   |||||||||||||||||||||||||   ||||||
    553 AAAGAAGAGCTCANAAGAAAGCCCAGATAGAAGAAGAGAAAANAAATGCA 602

    344 GluLysGluLysProGlnArgAsn 351
        ||||||||||   |||||||||
    603 GAAAAAGAAAAGCAGCAGAGAAAT 626

seq_name: gb_est1:AV227702

seq_documentation_block:
LOCUS       AV227702               988 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  AV227702 RIKEN full-length enriched, 14 days embryo liver Mus
            musculus cDNA clone 4432406P10 3', mRNA sequence.
ACCESSION   AV227702
VERSION     AV227702.2  GI:16385485
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1  (bases 1 to 988)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 1, 1999 this sequence version replaced gi:6179369.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .988
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="4432406P10"
                      /clone_lib="RIKEN full-length enriched, 14 days embryo
                      liver"
                      /sex="mixed"
                      /tissue_type="liver"
                      /dev_stage="14 days embryo"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
```

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 348 a 180 c 211 g 247 t 2 others
ORIGIN

alignment_scores:
Quality: 1072.00 Length: 326
Ratio: 3.815 Gaps: 0
Percent Similarity: 86.196 Percent Identity: 62.577

alignment_block:
US-09-836-410A-1 x AV227702 ..

Align seg 1/1 to: AV227702 from: 1 to: 988

```
193 CysSerLysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGl 209
    ||||||:::|||||||||||||||||   ||||||      ||||||||||
 11 TGCTCCAGTTTCACAAGAGAAGGACCATCTGCCATNGGGAATCTGAATGA 60

209 uMetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaM 226
    |   |||||||||   |||:::|||||||||   :::|||||||:::   :
 61 ATGACAGTGTATGTAGTTTGAGACGGAGTGCATCTCTGCCTATCAGCGTC 110

226 etAsnLysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHis 242
    ::::::::::   :::|||   |||:::|||||||||:::|||||||||
111 TGGGGAGATATGAGGATGCCTCGAAAGAGTGCCATGAAGTAGAGAGGCAT 160

243 PheIleGluIleThrAspAspGlnPheAspPheHisThrTyrCysMetAr 259
    |||   ||||||||||||||||||||||||||||||||||||||||||||
161 TTCTT.GAGATAACCGATGATCAGTTTGACTTCCATACATACTGCATGAG 209

259 gLysIleThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspValL 276
    ||||:::|||||||||:::||||||   ||||||:::|||||||||:::|
210 AAAGATGACCCTCCGTGCTTATGTTGGCCTCTTGAGATTAGAAGATGCTC 259

276 euArgGlnHisProPheTyrPheLysAlaAlaArgIleAlaIleGluIle 292
    ||||||:::|||   ||||||||||||||||||||   ||||||:::|||
260 TCAGAAGACATACTTTTTATTTCAAGGCTGCTAGATCAGCAATTGATATA 309

293 TyrLeuLysLeuHisAspAsnProLeuThrAspGluAsnLysGluHisGl 309
    ||||||||||||||||||||||||||||||:::::::::|||:::::::
310 TATTTGAAATTACATGATAACCCTTTAACCAATGACAGCAAACAACAAGA 359

309 uAlaAspThrAlaAsnMetSerAspLysGluLeuLysLysLeuArgAsnL 326
    :   |||:::   |||:::|||   |||||||:::|||||||:::   :::|
360 CATAGATTCAGAAAACCTGTCAGCCAAAGAAATGAAGAAAATGCTTAGCA 409

326 ysGlnArgArgAlaGlnLysLysAlaGlnIleGluGluGluLysLysAsn 342
    |||||||||||||||||||||||||||:::::|||||||||:::|||:::
410 AGCAAAGAAGAGCTCAGAAAAAGGCTAAGGTAGAAGAAGAGAGAAAGCAC 459

343 AlaGluLysGluLysProGlnArgAsnProLysLysLysLysAspAspAs 359
    :::|||:::|||:::   |||:::|||   |||||||||:::::::::::
460 ACAGAAAGGGAACGCCAACAGAAAAACCAAAAGAAAAAAAGAGAAGAAGA 509

359 pAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLysLeuA 376
    ::::|||      :::|||   ||||||||||||||||||||||||||
510 AGAAGAAGTTACCAGTGGTCATAAGGAAGAACTTATTCCTGAAAAACTAG 559

376 laLysValGluThrProLeuGluGluAlaIleLysPheLeuThrProLeu 392
      :::|||::::::||||||||||||||||||||||||||||||||||||
560 AAAGGGTTGACAATCCATTAGAAGAAGCCATCAAGTTCCTCACTCCTCTG 609

393 LysAsnLeuValLysAsnLysIleGluThrHisLeuPheAlaPheGluIl 409
    |||:::|||:::   :::::::|||:::||||||||||:::||||||||||
610 AAAACCCTTGCTGCTGAGAGCATTGATACCCACCTGCTGGCCTTTGAAAT 659

409 eTyrPheArgLysGluLysPheLeuLeuMetLeuGlnSerValLysArgA 426
    |||||||||||||   ||||||||||||||||||||||||||||||||||
660 ATATTTTAGAAAAGGGAAGTTTTTGCTAATGTTGCAGTCTGTTAAACGAG 709

426 laPheAlaIleAspSerSerHisProTrpLeuHisGluCysMetIleArg 442
    ||||||||||||:::|||:::::::||||||||||||||||||:::|||:::
710 CCTTTGCCATTGAAAGTAATAATCCTTGGTTACATGAATGCTTGATTAAG 759

443 LeuPheHisSerValCysGluSerLysAspLeuProGluThrValArgTh 459
    :::      ||||||   :::   ::::::||||||:::   |||
760 TTTTCTAAATCTGTCTCTAATCACAGTAATCTTCCAGACATTGTGAGCAA 809

459 rValLeuLysGlnGluMetAsnArgLeuPheGlyAlaThrAsnProLysA 476
    '||||||   ||||||||||:::::::::|||         :::      :
810 AGTTCTAGCTCAAGAAATGAAGAAGATATTTGTCAACAAAGATTTGCACA 859

476 snPheAsnGluThrPheLeuLysArgAsnSerAspSerLeuProHisArg 492
    ::||||||||||   ||||||::::::|||:::   ||||||   |||
860 GCTTTAATGAGGATTTCCTCAGACACAATGCTACCTCTCTTCAGCATCTA 909

493 LeuSerAlaAlaLysMetValTyrTyrLeuAspSerSerSerGlnLysAr 509
    |||:::::::||||||||:::|||:::||||||:::|||   |||:::::
910 CTTGCAGGTGCTAAAATGATGTATTTTCTGGACAAGTCAAGGCAAGAAAA 959

509 gAlaIleGluLeuAlaThrThrLeuAsp 518
    :||||||      ||||||   ||||||
960 AGCAATTGCTACAGCCACTAGGTTGGAT 987
```

seq_name: gb_est2:BJ057588

seq_documentation_block:

LOCUS BJ057588 638 bp mRNA linear EST 10-DEC-2001
DEFINITION BJ057588 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL104i04 5', mRNA sequence.
ACCESSION BJ057588
VERSION BJ057588.1 GI:17479669
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 638)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .638
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL104i04"
/clone_lib="NIBB Mochii normalized Xenopus tailbud library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
BASE COUNT 257 a 98 c 151 g 132 t
ORIGIN

alignment_scores:
Quality: 1071.00 Length: 208
Ratio: 5.199 Gaps: 0
Percent Similarity: 99.038 Percent Identity: 96.635

alignment_block:
US-09-836-410A-1 x BJ057588 ..

Align seg 1/1 to: BJ057588 from: 1 to: 638

```
157 ArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAspArgPheIleAs 173
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 15 CGATGGATGGATGAGGCGCAGGCACTGGATACAGCTGACAGATTTATCAA 64

173 nSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIleLysGluAlaG 190
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 65 TTCTAAATGTGCAAAATATATGTTAAAAGCAAACCTGATTAAAGAGGCTG 114

190 luGluMetCysSerLysPheThrArgGluGlyThrSerAlaValGluAsn 206
    ||||||||||||||||||||||||||||||||||||||||||||||||||
115 AAGAAATGTGCTCGAAATTTACAAGGGAAGGAACATCAGCAGTGGAAAAT 164

207 LeuAsnGluMetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTy 223
    ||||||||||||||||||||||||||||||||||||||||||||||||||
165 CTGAACGAGATGCAGTGCATGTGGTTCCAGACAGAATGTGCACAAGCTTA 214

223 rLysAlaMetAsnLysPheGlyGluAlaLeuLysLysCysHisGluIleG 240
    ||||:::|||||||||:::||||||||||||||||||||||||||||||||
215 CAAATCCATGAATAAATATGGCGAAGCACTTAAAAAATGCCATGAAATTG 264

240 luArgHisPheIleGluIleThrAspAspGlnPheAspPheHisThrTyr 256
    ||||||||||||:::|||||||||||||||||||||||||||||||||||
265 AAAGGCATTTTGTAGAAATAACAGATGACCAGTTTGATTTCCACACTTAC 314

257 CysMetArgLysIleThrLeuArgSerTyrValAspLeuLeuLysLeuGl 273
    ||||||||||||||||||||||||||||||||||||||||||||||||||
315 TGTATGAGAAAAATTACACTCAGGTCATATGTGGACTTGTTAAAATTAGA 364

273 uAspValLeuArgGlnHisProPheTyrPheLysAlaAlaArgIleAlaI 290
    ||||||||||||||||||||||||||||||||||||||||||||||||||
365 GGATGTACTAAGGCAGCATCCATTTTACTTCAAGGCTGCACGGATTGCAA 414

290 leGluIleTyrLeuLysLeuHisAspAsnProLeuThrAspGluAsnLys 306
    ||||||||||||||||||||||||||||||||||||||||||||||||||
415 TAGAAATTTATCTGAAACTTCATGATAATCCTTTAACAGATGAGAATAAG 464

307 GluHisGluAlaAspThrAlaAsnMetSerAspLysGluLeuLysLysLe 323
    ||||||||||||||||||||||||||||||||||||||||||||||||||
465 GAGCATGAGGCAGACACAGCCAATATGTCTGACAAAGAGCTAAAGAAACT 514

323 uArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGluGluL 340
    ||||||||||||||||||||||||||||||||||:::||||||||||||
515 ACGTAACAAACAACGGAGGGCGCAGAAGAAAGCACAGCTGGAAGAGGAGA 564

340 ysLysAsnAlaGluLysGluLysProGlnArgAsnProLysLysLysLys 356
    ||||||||||||||||||||||   |||||||||   ||||||||||||
565 AAAAGAATGCAGAAAAAGAAAAGCAACAGAGAAACCAAAAGAAGAAAAAG 614

357 AspAspAspAspGluGluIleGly 364
    :::|||||||||||||||||||||
615 GAAGATGATGATGAAGAGATTGGA 638
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 19, 2002, 22:39:02 ; Search time 4131.3 Seconds
(without alignments)
17313.413 Million cell updates/sec

Title: US-09-836-410A-2
Perfect score: 3418
Sequence: 1 caagtaacacccgcaagatg.....atgcaataaattgttttggg 3418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1.	3418	100.0	3418	6	AX285242	AX285242 Sequence
	2	3418	100.0	3418	6	AX285294	AX285294 Sequence
c	3	3418	100.0	3418	6	AX285296	AX285296 Sequence
	4	3387	99.1	3421	10	AF237622	AF237622 Mus muscu
	5	2740.2	80.2	5505	9	HSA314788	AJ314788 Homo sapi
	6	2543	74.4	4192	9	AF327722	AF327722 Homo sapi
	7	1779	52.0	1779	6	AX285247	AX285247 Sequence
	8	1469.4	43.0	3324	5	AF247679	AF247679 Xenopus l
c	9	1413	41.3	1413	6	AX285295	AX285295 Sequence
	10	1210.2	35.4	1802	9	AK023387	AK023387 Homo sapi
c	11	1191.8	34.9	101276	2	AC094440	AC094440 Rattus no
	12	1126	32.9	2859	9	AK001595	AK001595 Homo sapi
	13	1079.6	31.6	1985	9	AK023402	AK023402 Homo sapi
	14	1038.4	30.4	145395	9	AC097376	AC097376 Homo sapi
	15	626.6	18.3	66729	2	AC102860	AC102860 Mus muscu
c	16	617.8	18.1	710	6	AX210531	AX210531 Sequence
	17	448.2	13.1	843	9	BC017099	BC017099 Homo sapi
	18	439	12.8	2917	9	AB049844	AB049844 Macaca fa
c	19	419	12.3	181573	2	AC020959	AC020959 Mus muscu
	20	397.8	11.6	1876	9	AB070089	AB070089 Macaca fa
	21	370.6	10.8	2402	9	AK025707	AK025707 Homo sapi
	22	342.6	10.0	408	6	AX245855	AX245855 Sequence
	23	340	9.9	408	6	AX245854	AX245854 Sequence
	24	339.8	9.9	402	6	AX245478	AX245478 Sequence
	25	281	8.2	404	6	AX071730	AX071730 Sequence
c	26	239	7.0	255	6	AX210853	AX210853 Sequence
	27	229.8	6.7	26121	2	AC015346	AC015346 Drosophil
	28	229.8	6.7	180213	3	AC010847	AC010847 Drosophil
	29	229.8	6.7	182623	3	AC011071	AC011071 Drosophil
	30	229.8	6.7	301457	3	AE003512	AE003512 Drosophil
	31	226	6.6	288	6	AX210600	AX210600 Sequence
	32	215.4	6.3	255	6	AX210898	AX210898 Sequence
c	33	212.6	6.2	66451	2	AC105966	AC105966 Mus muscu
	34	211.4	6.2	3295	8	AY064015	AY064015 Arabidops
	35	194.6	5.7	66451	2	AC105966	AC105966 Mus muscu
	36	150	4.4	38734	8	SPCC338	AL023781 S.pombe c
	37	119	3.5	101276	2	AC094440	AC094440 Rattus no
	38	118.8	3.5	122450	3	AC024801	AC024801 Caenorhab
	39	118.8	3.5	291622	2	AC006796	AC006796 Caenorhab
	40	116.6	3.4	54757	9	AL590665	AL590665 Human DNA
c	41	105.6	3.1	66729	2	AC102860	AC102860 Mus muscu
	42	78.6	2.3	109900	8	AF263282	AF263282 Filobasid
c	43	75	2.2	7218	6	I66494	I66494 Sequence 14
	44	73.4	2.1	2698	8	YSCNACT	M23166 S.cerevisia
	45	73.4	2.1	2699	6	I08122	I08122 Sequence 1

ALIGNMENTS

RESULT 1
AX285242
LOCUS AX285242 3418 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 1 from Patent WO0179506.
ACCESSION AX285242
VERSION AX285242.1 GI:17045930
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Gendron,R.L. and Paradis,H.
TITLE Treatment of ocular neovascularization and related diseases
JOURNAL Patent: WO 0179506-A 1 25-OCT-2001;
Children's Hospital Research Foundation (US)
FEATURES Location/Qualifiers
source 1. .3418
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1157 a 604 c 704 g 953 t
ORIGIN

Query Match 100.0%; Score 3418; DB 6; Length 3418;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db       1 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 60

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db      61 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 120

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db     121 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 180

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
Db     181 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 240

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
Db     241 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 300

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
Db     301 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 360

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
Db     361 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 420

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
Db     421 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 480

Qy     481 ccttaaacttttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
Db     481 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 540

Qy     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
Db     541 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 600

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
Db     601 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 660

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
Db     661 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 720

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
Db     721 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 780

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
Db     781 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 840

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
Db     841 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 900

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
Db     901 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 960

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
Db     961 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 1020

Qy    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db    1021 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 1080

Qy    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db    1081 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 1140

Qy    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db    1141 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 1200

Qy    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccattttacttcaaag 1260
Db    1201 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 1260

Qy    1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Db    1261 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 1320

Qy    1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Db    1321 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 1380

Qy    1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Db    1381 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAAAAAAATGCCGAAA 1440

Qy    1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
Db    1441 AAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAG 1500

Qy    1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Db    1501 GCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAG 1560

Qy    1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Db    1561 CTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTT 1620

Qy    1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Db    1621 TTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGC 1680

Qy    1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Db    1681 GGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTC 1740

Qy    1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Db    1741 ATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAA 1800

Qy    1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttttctgaaaagga 1860
Db    1801 TGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGA 1860

Qy    1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Db    1861 ATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTT 1920

Qy    1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Db    1921 CTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTCACCAACAGAA 1980

Qy    1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Db    1981 ACCTTCAGACTTGCATGGAAGTGTTGGAAGCCTTGTGTGATGGTAGCCTACGAGACTGTA 2040

Qy    2041 aagaagctgccgaagcctacagagcaagttgtcataagctttttcccttatgctttggctt 2100
Db    2041 AAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTT 2100

Qy    2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
```

```
Db    2101 TCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTGCAG 2160

Qy    2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Db    2161 AAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGAATGACTTT 2220

Qy    2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Db    2221 GGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTACAC 2280

Qy    2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgtttttttatcctgctgaa 2340
Db    2281 AGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATCCTGCTGAA 2340

Qy    2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Db    2341 AACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTAAAA 2400

Qy    2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Db    2401 GCTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTCCCT 2460

Qy    2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Db    2461 TCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAACATAGCAGGT 2520

Qy    2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
Db    2521 GTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCTCAGACAGCA 2580

Qy    2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Db    2581 TAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTT 2640

Qy    2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
Db    2641 TAGCATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATGTGA 2700

Qy    2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
Db    2701 ATCACTGTGTACAGAGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACACATA 2760

Qy    2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
Db    2761 GCTACCAAGCGGATTAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAA 2820

Qy    2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
Db    2821 ATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAG 2880

Qy    2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
Db    2881 TTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTTTAT 2940

Qy    2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
Db    2941 CATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGG 3000

Qy    3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
Db    3001 CACCAACTAATTATGCAAGCATGAATTTTTCACCTGAGAGTGAAAAAAGAAAACTCTACC 3060

Qy    3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
Db    3061 ATGGCTTGAAGTTACAGGAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAGACT 3120

Qy    3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
Db    3121 AATATCTAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTTAGA 3180

Qy    3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
Db    3181 TACCTCTGGGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTCT 3240

Qy    3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
Db    3241 AAGAATATATTTCTCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGG 3300

Qy    3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360
Db    3301 GCCCAAGATTTTTTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGTAAT 3360

Qy    3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
Db    3361 TTCTTTTTTTTTTTGTTGTCAAACATAAGGTACCAAATATGCAATAAATTGTTTTGGG 3418


RESULT   2
AX285294
LOCUS       AX285294                3418 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 2 from Patent WO0179505.
ACCESSION   AX285294
VERSION     AX285294.1  GI:17045975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 2 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .3418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1157 a     604 c     704 g     953 t
ORIGIN


  Query Match           100.0%;  Score 3418;  DB 6;  Length 3418;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db       1 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 60

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db      61 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 120

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db     121 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 180

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
Db     181 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 240

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
Db     241 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 300

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
Db     301 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 360

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
Db     361 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 420

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
```

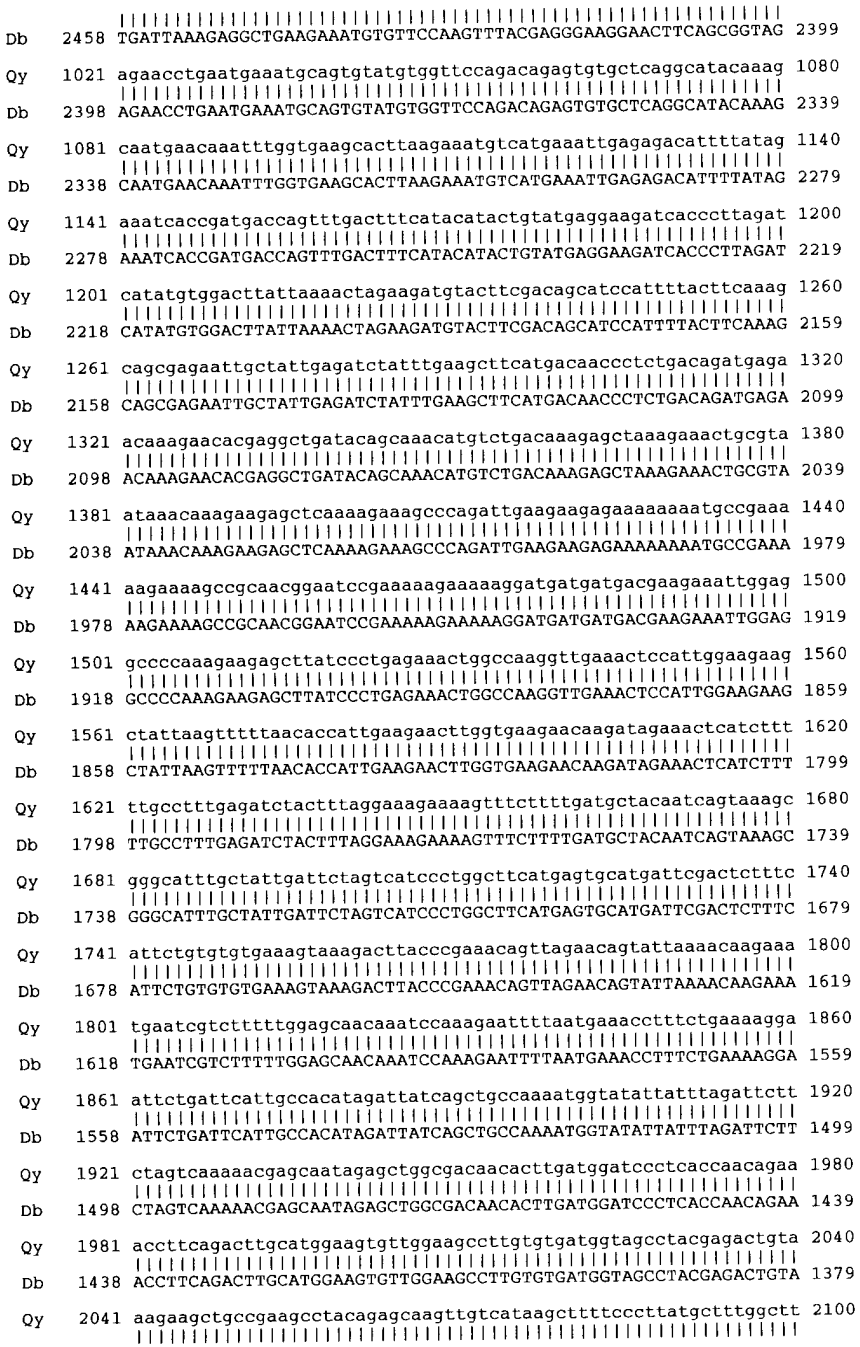
```
Db    421 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 480
Qy    481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
Db    481 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 540
Qy    541 atttcagcaagggctgtccacctgtcttcaatacctgaggtctttatacagagataaag 600
Db    541 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 600
Qy    601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
Db    601 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 660
Qy    661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
Db    661 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 720
Qy    721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
Db    721 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 780
Qy    781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
Db    781 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 840
Qy    841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
Db    841 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 900
Qy    901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
Db    901 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 960
Qy    961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
Db    961 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 1020
Qy   1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db   1021 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 1080
Qy   1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db   1081 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 1140
Qy   1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db   1141 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 1200
Qy   1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccattttacttcaaag 1260
Db   1201 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 1260
Qy   1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Db   1261 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 1320
Qy   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Db   1321 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 1380
Qy   1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Db   1381 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAAAAAAATGCCGAAA 1440
Qy   1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
Db   1441 AAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAG 1500
Qy   1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Db   1501 GCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAG 1560
Qy   1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Db   1561 CTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTT 1620
Qy   1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Db   1621 TTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGC 1680
Qy   1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Db   1681 GGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTC 1740
Qy   1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Db   1741 ATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAA 1800
Qy   1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaacctttctgaaaagga 1860
Db   1801 TGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGA 1860
Qy   1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Db   1861 ATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTT 1920
Qy   1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Db   1921 CTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTCACCAACAGAA 1980
Qy   1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Db   1981 ACCTTCAGACTTGCATGGAAGTGTTGGAAGCCTTGTGTGATGGTAGCCTACGAGACTGTA 2040
Qy   2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100
Db   2041 AAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTT 2100
Qy   2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
Db   2101 TCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTGCAG 2160
Qy   2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Db   2161 AAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGAATGACTTT 2220
Qy   2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Db   2221 GGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTACAC 2280
Qy   2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgctgaa 2340
Db   2281 AGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATCCTGCTGAA 2340
Qy   2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Db   2341 AACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTAAAA 2400
Qy   2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Db   2401 GCTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTCCCT 2460
Qy   2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Db   2461 TCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAACATAGCAGGT 2520
Qy   2521 gtgtggcagaaatattacttttaaatttgtcttgtgagattttgctatatctcagacagca 2580
Db   2521 GTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCTCAGACAGCA 2580
Qy   2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Db   2581 TAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTT 2640
```

```
Qy   2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TAGCATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATGTGA 2700

Qy   2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 ATCACTGTGTACAGAGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACACATA 2760

Qy   2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCTACCAAGCGGATTAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAA 2820

Qy   2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 ATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAG 2880

Qy   2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTTTAT 2940

Qy   2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGG 3000

Qy   3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CACCAACTAATTATGCAAGCATGAATTTTTCACCTGAGAGTGAAAAAAGAAAACTCTACC 3060

Qy   3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATGGCTTGAAGTTACAGGAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAGACT 3120

Qy   3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AATATCTAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTTAGA 3180

Qy   3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 TACCTCTGGGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTCT 3240

Qy   3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 AAGAATATATTTCTCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGG 3300

Qy   3301 gcccaagatttttttttcaataacagaccagcttcttttttcctgcagttacaaatgtaat 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GCCCAAGATTTTTTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGTAAT 3360

Qy   3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TTCTTTTTTTTTTTGTTGTCAAACATAAGGTACCAAATATGCAATAAATTGTTTTGGG 3418
```

```
RESULT   3
AX285296/c
LOCUS       AX285296                3418 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 4 from Patent WO0179505.
ACCESSION   AX285296
VERSION     AX285296.1  GI:17045977
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 4 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .3418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
BASE COUNT      953 a    704 c    604 g   1157 t
ORIGIN
```

```
  Query Match           100.0%;  Score 3418;  DB 6;  Length 3418;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3418 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 3359

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaatttttagaagagtttaggaaaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3358 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 3299

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3298 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 3239

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3238 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 3179

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3178 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 3119

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3118 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 3059

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3058 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 2999

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2998 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 2939

Qy     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2938 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 2879

Qy     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2878 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 2819

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2818 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 2759

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2758 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 2699

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2698 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 2639

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2638 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 2579

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2578 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 2519

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2518 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 2459

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
```


Db 2458 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 2399
Qy 1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db 2398 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 2339
Qy 1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db 2338 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 2279
Qy 1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db 2278 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 2219
Qy 1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260
Db 2218 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 2159
Qy 1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Db 2158 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 2099
Qy 1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Db 2098 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 2039
Qy 1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Db 2038 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAAAAAAATGCCGAAA 1979
Qy 1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
Db 1978 AAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAG 1919
Qy 1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Db 1918 GCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAG 1859
Qy 1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Db 1858 CTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTT 1799
Qy 1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Db 1798 TTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGC 1739
Qy 1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Db 1738 GGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTC 1679
Qy 1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Db 1678 ATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAA 1619
Qy 1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaacctttctgaaaagga 1860
Db 1618 TGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGA 1559
Qy 1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Db 1558 ATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTT 1499
Qy 1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Db 1498 CTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTCACCAACAGAA 1439
Qy 1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Db 1438 ACCTTCAGACTTGCATGGAAGTGTTGGAAGCCTTGTGTGATGGTAGCCTACGAGACTGTA 1379
Qy 2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100
Db 1378 AAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTT 1319
Qy 2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
Db 1318 TCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTGCAG 1259
Qy 2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Db 1258 AAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGAATGACTTT 1199
Qy 2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Db 1198 GGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTACAC 1139
Qy 2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgctgaa 2340
Db 1138 AGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATCCTGCTGAA 1079
Qy 2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Db 1078 AACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTAAAA 1019
Qy 2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Db 1018 GCTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTCCCT 959
Qy 2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Db 958 TCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAACATAGCAGGT 899
Qy 2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
Db 898 GTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCTCAGACAGCA 839
Qy 2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Db 838 TAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTT 779
Qy 2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
Db 778 TAGCATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATGTGA 719
Qy 2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
Db 718 ATCACTGTGTACAGAGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACACATA 659
Qy 2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
Db 658 GCTACCAAGCGGATTAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAA 599
Qy 2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
Db 598 ATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAG 539
Qy 2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
Db 538 TTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTTTAT 479
Qy 2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
Db 478 CATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGG 419
Qy 3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
Db 418 CACCAACTAATTATGCAAGCATGAATTTTTCACCTGAGAGTGAAAAAAGAAAACTCTACC 359
Qy 3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
Db 358 ATGGCTTGAAGTTACAGGAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAGACT 299
Qy 3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
Db 298 AATATCTAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTTAGA 239

```
Qy   3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 TACCTCTGGGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTCT 179

Qy   3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaaagg 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 AAGAATATATTTCTCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGG 119

Qy   3301 gcccaagatttttttttcaataacagaccagcttcttttttcctgcagttacaaatgtaat 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GCCCAAGATTTTTTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGTAAT 59

Qy   3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     58 TTCTTTTTTTTTTTGTTGTCAAACATAAGGTACCAAATATGCAATAAATTGTTTTGGG 1
```

```
RESULT   4
AF237622
LOCUS       AF237622                3421 bp    mRNA    linear   ROD 13-JUN-2000
DEFINITION  Mus musculus acetyltransferase Tubedown-1 mRNA, complete cds.
ACCESSION   AF237622
VERSION     AF237622.1  GI:8164012
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Tubedown-1, A novel acetyltransferase associated with blood vessel
            development
  JOURNAL   Dev. Dyn. 218 (2), 300-315 (2000)
  MEDLINE   20302532
REFERENCE   2  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Pediatrics, Childrens Hospital Medical
            Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA
REFERENCE   3  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Pediatrics, Childrens Hospital Medical
            Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA
  REMARK    Amino acid sequence updated by submitter
FEATURES             Location/Qualifiers
     source          1. .3421
                     /organism="Mus musculus"
                     /strain="ICR outbred"
                     /db_xref="taxon:10090"
     CDS             408. .2192
                     /codon_start=1
                     /product="acetyltransferase Tubedown-1"
                     /protein_id="AAF73953.2"
                     /db_xref="GI:8497318"
                     /translation="MLERLKIYEEAWTKYPRGLVPRKLPLNFLSGEKFKECLDRFLRM
                     NFSKGCPPVFNTLRSLYRDKEKVAIVEELVVGYETSLKSCRLFNPNDDGKEEPPTTLL
                     WVQYYLAQHYDKIGQPSIALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMD
                     EAQALDTADRFINSKCAKYMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTE
                     CAQAYKAMNKFGEALKKCHEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVL
                     RQHPFYFKAARIAIEIYLKLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKA
                     QIEEEKKNAEKEKPQRNPKKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPL
                     KNLVKNKIETHLFAFEIYFRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFHSVCES
                     KDLPETVRTVLKQEMNRLFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDSSSQK
                     RAIELATTLDGIPHQQKPSDLHGSVGKPCCDGSLRDCKEAAEAYRASCHKLFPYALAF
                     MPPGYEEDMKITVNGDSSAETEELANEI"
BASE COUNT     1157 a    604 c    706 g    954 t
ORIGIN
```

```
  Query Match           99.1%;  Score 3387;  DB 10;  Length 3421;
  Best Local Similarity 99.8%;  Pred. No. 0;
```

```
  Matches 3413;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    2;

Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 60

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaatttttagaagagtttaggaaaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 120

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 180

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 240

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 300

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 360

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 420

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 480

Qy     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCTTAAACTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 540

Qy     541 atttcagcaagggctgtccacctgtcttcaatacctgaggtctttatacagagataaag 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 600

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 660

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 720

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 780

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 840

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 900

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 960

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 1020

Qy    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 1080
```

```
Qy   1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacatttttatag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 1140

Qy   1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 1200

Qy   1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 1260

Qy   1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 1320

Qy   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 1380

Qy   1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAAAAAAATGCCGAAA 1440

Qy   1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAG 1500

Qy   1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAG 1560

Qy   1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTT 1620

Qy   1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGC 1680

Qy   1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTC 1740

Qy   1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAA 1800

Qy   1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaacctttctgaaaagga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGA 1860

Qy   1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTT 1920

Qy   1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgat-ggatccctcaccaacaga 1979
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1921 CTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGGATCCCTCACCAACAGA 1980

Qy   1980 aaccttcagacttgcatggaagtgttgg--aagccttgtgtgatggtagcctacgagact 2037
          ||||||||||||||||||||||||||  | | | ||||||||||||||||||||||||||
Db   1981 AACCTTCAGACTTGCATGGAAGTGTTGGGAAGCCTTGTTGTGATGGTAGCCTACGAGACT 2040

Qy   2038 gtaaagaagctgccgaagcctacagagcaagttgtcataagctttttcccttatgctttgg 2097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GTAAAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGG 2100

Qy   2098 ctttcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctg 2157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CTTTCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 2160
```

```
Qy   2158 cagaaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgac 2217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CAGAAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGAATGAC 2220

Qy   2218 tttggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttactta 2277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TTTGGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTA 2280

Qy   2278 cacagaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgct 2337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CACAGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATCCTGCT 2340

Qy   2338 gaaaactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaatta 2397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GAAAACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTA 2400

Qy   2398 aaagctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttc 2457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AAAGCTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTC 2460

Qy   2458 ccttcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagca 2517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 CCTTCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAACATAGCA 2520

Qy   2518 ggtgtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagaca 2577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GGTGTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCTCAGACA 2580

Qy   2578 gcataaataaatgctgttttagcactggattctttcactgagcacaaagagttgttgggg 2637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GCATAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGG 2640

Qy   2638 ctttagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatg 2697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTTTAGCATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATG 2700

Qy   2698 tgaatcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacac 2757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TGAATCACTGTGTACAGAGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACAC 2760

Qy   2758 atagctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttcc 2817
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATAGCTACCAAGCGGATTAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCC 2820

Qy   2818 aaaattcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaac 2877
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 AAAATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAAC 2880

Qy   2878 cagttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctt 2937
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CAGTTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTT 2940

Qy   2938 tatcattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatg 2997
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TATCATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATG 3000

Qy   2998 tggcaccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactct 3057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 TGGCACCAACTAATTATGCAAGCATGAATTTTTCACCTGAGAGTGAAAAAGAAAACTCT 3060

Qy   3058 accatggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagag 3117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ACCATGGCTTGAAGTTACAGGAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAG 3120

Qy   3118 actaatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagttt 3177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 ACTAATATCTAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTT 3180

Qy   3178 agatacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatt 3237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGATACCTCTGGGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATT 3240

Qy   3238 tctaagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaa 3297
```

```
Db    3241 TCTAAGAATATATTTCTCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAA 3300

Qy    3298 agggcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgt 3357

Db    3301 AGGGCCCAAGATTTTTTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGT 3360

Qy    3358 aatttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttgg 3417

Db    3361 AATTTCTTTTTTTTTTTGTTGTCAAACATAAGGTACCAAATATGCAATAAATTGTTTTGG 3420

Qy    3418 g 3418

Db    3421 G 3421
```

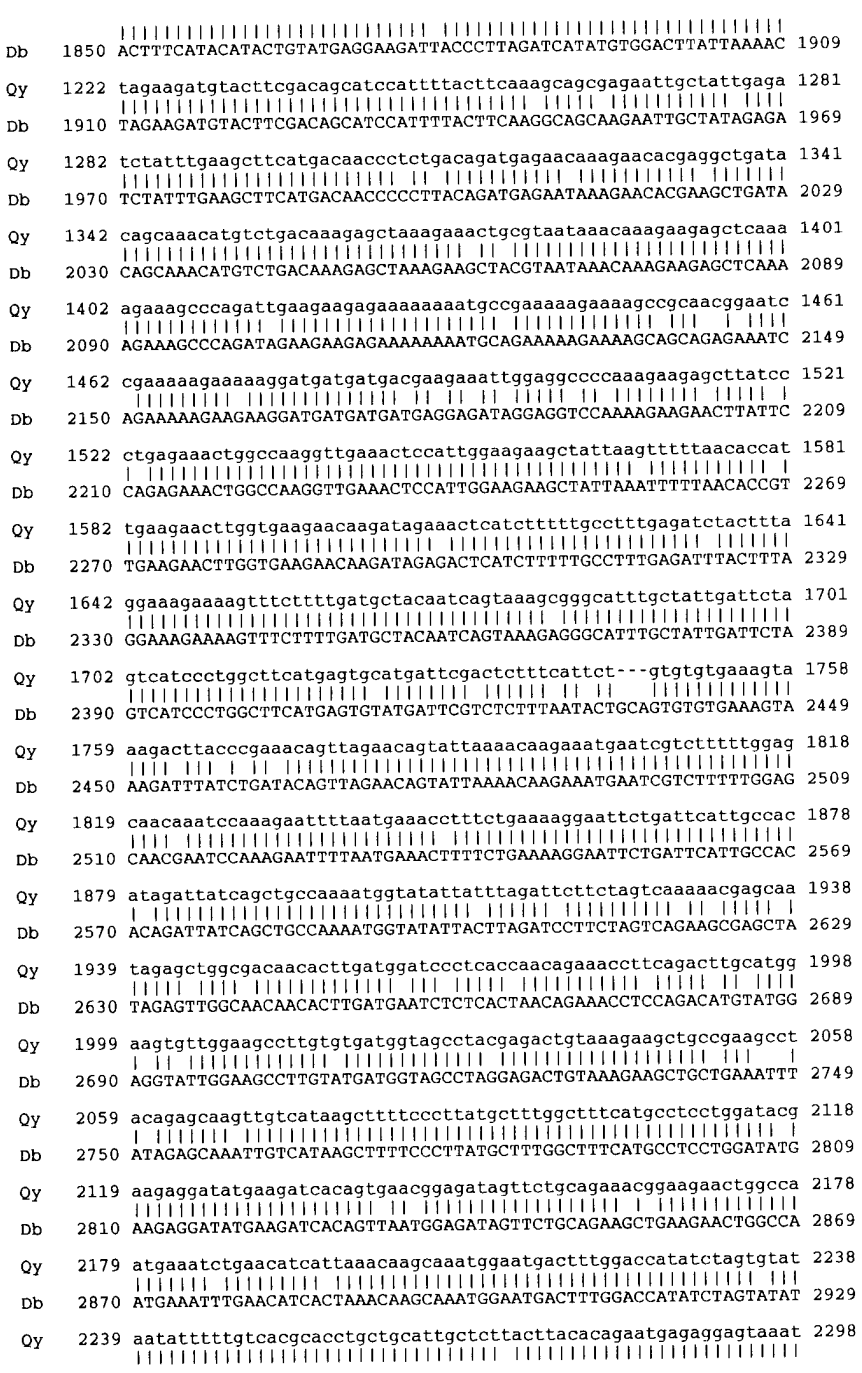
```
RESULT   5
HSA314788
LOCUS       HSA314788               5505 bp    mRNA    linear   PRI 01-JUL-2001
DEFINITION  Homo sapiens mRNA for putative N-acetyltransferase.
ACCESSION   AJ314788
VERSION     AJ314788.1  GI:14589341
KEYWORDS    N-acetyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5505)
  AUTHORS   Fluge,O., Bruland,O., Akslen,L.A., Varhaug,J.E. and Lillehaug,J.R.
  TITLE     Identification of NATH, a novel gene overexpressed in papillary
            thyroid carcinomas
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5505)
  AUTHORS   Fluge,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2001) Fluge O., Dept. of Molecular Biology,
            University of Bergen, Thormohlens gt 55, N-5020 Bergen, NORWAY
COMMENT     related entry AF327722.
FEATURES             Location/Qualifiers
     source          1. .5505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="papillary thyroid carcinoma"
     CDS             209. .2809
                     /codon_start=1
                     /evidence=experimental
                     /product="putative N-acetyltransferase"
                     /protein_id="CAC43228.1"
                     /db_xref="GI:14589342"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                     LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                     QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                     LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                     DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                     ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                     YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                     HEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVLRQHPFYFKAARIAIEIYL
                     KLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKAQIEEEKKNAEKEKQQRNQ
                     KKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPLKNLVKNKIETHLFAFEIY
                     FRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFNTAVCESKDLSDTVRTVLKQEMNR
                     LFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDPSSQKRAIELATTLDESLTNRN
                     LQTCMEVLEALYDGSLGDCKEAAEIYRANCHKLFPYALAFMPPGYEEDMKITVNGDSS
                     AEAEELANEI"
     conflict        3752. .5505
                     /citation=[1]
BASE COUNT     1847 a    914 c   1154 g   1590 t
ORIGIN
```

```
Query Match           80.2%;  Score 2740.2;  DB 9;  Length 5505;
Best Local Similarity 91.4%;  Pred. No. 0;
```

```
  Matches 3129;  Conservative    0;  Mismatches  238;  Indels   56;  Gaps   19;

Qy      27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81

Db     639 GACCTGCGCAGAGAGCATCATGGATTGGTTATGCTATTGCTTACCATTTATTAGAAGATT 698

Qy      82 atgaaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgata 141

Db     699 ATGAAATGGCAGCAAAGATTTTAGAAGAATTTAGGAAAACACAACAGACATCCCCTGACA 758

Qy     142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201

Db     759 AGGTGGATTATGAATATAGTGAACTACTCTTATATCAGAATCAAGTTCTTCGGGAAGCAG 818

Qy     202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261

Db     819 GTCTCTATAGAGAAGCTTTGGAACATCTTTGTACCTATGAAAAGCAGATTTGTGATAAAC 878

Qy     262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321

Db     879 TTGCTGTAGAAGAAACCAAAGGGGAACTTCTGTTGCAACTATGTCGTTTGGAAGATGCTG 938

Qy     322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381

Db     939 CAGATGTTTATAGAGGATTGCAAGAGAGAAATCCTGAAAACTGGGCCTATTACAAAGGCT 998

Qy     382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441

Db     999 TGGAAAAAGCACTCAAGCCAGCTAATATGTTAGAACGGCTAAAAATTTATGAGGAAGCCT 1058

Qy     442 ggactaaatacccccaggggactcgtgccaagaaagctgcccttaaacttttttatctggag 501

Db    1059 GGACTAAATATCCCAGGGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTG 1118

Qy     502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561

Db    1119 AGAAGTTTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGCCCAC 1178

Qy     562 ctgtcttcaataccttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621

Db    1179 CAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGTGGCAATCATAGAAG 1238

Qy     622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681

Db    1239 AGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATG 1298

Qy     682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741

Db    1299 ATGGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATT 1358

Qy     742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801

Db    1359 ATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTA 1418

Qy     802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861

Db    1419 CACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATA 1478

Qy     862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921

Db    1479 TTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTA 1538

Qy     922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981

Db    1539 TCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAA 1598

Qy     982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041

Db    1599 TGTGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGT 1658

Qy    1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101

Db    1659 GCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAG 1718
```

```
Qy   1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
Db   1719 CACTTAAGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTG 1778

Qy   1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
Db   1779 ACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTTATTAAAAC 1838

Qy   1222 tagaagatgtacttcgacagcatccatttacttcaaagcagcgagaattgctattgaga 1281
Db   1839 TAGAAGATGTACTTCGACAGCATCCATTTACTTCAAGGCAGCAAGAATTGCTATAGAGA 1898

Qy   1282 tctatttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgata 1341
Db   1899 TCTATTTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGCTGATA 1958

Qy   1342 cagcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaa 1401
Db   1959 CAGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAA 2018

Qy   1402 agaaagcccagattgaagaagagaaaaaaaatgccgaaaaagaaaagccgcaacggaatc 1461
Db   2019 AGAAAGCCCAGATAGAAGAAGAGAAAAAAAATGCAGAAAAAGAAAAGCAGCAGAGAAATC 2078

Qy   1462 cgaaaaagaaaaaggatgatgatgacgaagaaattggaggccccaaagaagagcttatcc 1521
Db   2079 AGAAAAAGAAGAAGGATGATGATGATGAGGAGATAGGAGGTCCAAAAGAAGAACTTATTC 2138

Qy   1522 ctgagaaactggccaaggttgaaactccattggaagaagctattaagtttttaacaccat 1581
Db   2139 CAGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAGCTATTAAATTTTTAACACCGT 2198

Qy   1582 tgaagaacttggtgaagaacaagatagaaactcatcttttgcctttgagatctactta 1641
Db   2199 TGAAGAACTTGGTGAAGAACAAGATAGAGACTCATCTTTTTGCCTTTGAGATTTACTTA 2258

Qy   1642 ggaaagaaaagtttctttttgatgctacaatcagtaaagcgggcatttgctattgattcta 1701
Db   2259 GGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGAGGGCATTTGCTATTGATTCTA 2318

Qy   1702 gtcatccctggcttcatgagtgcatgattcgactctttcattct---gtgtgtgaaagta 1758
Db   2319 GTCATCCCTGGCTTCATGAGTGTATGATTCGTCTCTTTAATACTGCAGTGTGTGAAAGTA 2378

Qy   1759 aagacttacccgaaacagttagaacagtattaaaacaagaaatgaatcgtcttttggag 1818
Db   2379 AAGATTTATCTGATACAGTTAGAACAGTATTAAAACAAGAAATGAATCGTCTTTTTGGAG 2438

Qy   1819 caacaaatccaaagaattttaatgaaacctttctgaaaaggaattctgattcattgccac 1878
Db   2439 CAACGAATCCAAAGAATTTTAATGAAACTTTTCTGAAAAGGAATTCTGATTCATTGCCAC 2498

Qy   1879 atagattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaa 1938
Db   2499 ACAGATTATCAGCTGCCAAAATGGTATATTACTTAGATCCTTCTAGTCAGAAGCGAGCTA 2558

Qy   1939 tagagctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatgg 1998
Db   2559 TAGAGTTGGCAACAACACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGG 2618

Qy   1999 aagtgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcct 2058
Db   2619 AGGTATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAAATTT 2678

Qy   2059 acagagcaagttgtcataagcttttccccttatgctttggctttcatgcctcctggatacg 2118
Db   2679 ATAGAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCCTCCTGGATATG 2738

Qy   2119 aagaggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggcca 2178
Db   2739 AAGAGGATATGAAGATCACAGTTAATGGAGATAGTTCTGCAGAAGCTGAAGAACTGGCCA 2798

Qy   2179 atgaaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtat 2238
Db   2799 ATGAAATTTGAACATCACTAAACAAGCAAATGGAATGACTTTGGACCATATCTAGTATAT 2858

Qy   2239 aatattttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaat 2298
Db   2859 AATATTTTTGTCACGCACCTGCTGCATTGCTCTAACTTACACAGAATGAGAGGAGTAAAT 2918

Qy   2299 gttcttgccttcaaata--gtcttacgtttttttatcctgctg-aaaactatatataaaat 2355
Db   2919 GTTCTTGCCTTCAAATAGTGTTTTACGTTTTTTATCCTGCTGAAAAAGTATATATAAAAT 2978

Qy   2356 atctaacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgag 2415
Db   2979 ATCTAACATTACAGGATAGAGGTTCAGTTTCTTAAAAAATT-AAAGCTGCTAAAATTGAG 3037

Qy   2416 gggtttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaac 2475
Db   3038 TGGTT------AAAAAAGATACCTTATCCTATTCCTCCCCACCCACCCATGTTTTTAAAC 3091

Qy   2476 taatttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatat 2535
Db   3092 TAATTTATATAAAATCTGGAGGCTGTTACAGCTAACAAAGCAGGTGTGTGGCAGAAATAT 3151

Qy   2536 tactttaaattgtcttgtgagatttttgctatatctcagacagcataaataaatgctgtt 2595
Db   3152 TACTTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC----ATAAATGCTGTT 3207

Qy   2596 ttagcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgat 2655
Db   3208 TTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTTTAGCATCTGACTGAT 3267

Qy   2656 tctgttacggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacaga 2715
Db   3268 TTTGTTACGGGGTTGAT-TCTGACCATAGGAAGTATGCAATGTGAATCACTATTTACAGA 3326

Qy   2716 gccgtctacaacacatgcttgacgttgtagagactgggacacatagctaccaagc-ggat 2774
Db   3327 GAAACCTACAACAGATGCTTGATGTTGTAGAAACTGGGACATATAGATACCAAGCAAAAT 3386

Qy   2775 taagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatg 2834
Db   3387 TATAAGAAACCTATAAGGTGTTCAATACG---CTTGTGTTTCCAAAATTCACTGTACATG 3443

Qy   2835 atcagtttggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctca 2894
Db   3444 ATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACTGAAGGAACCAGTTGTAACAGTCTCA 3503

Qy   2895 a-tttaactaaaacttgaagaactaaaataacaatgcaaacctttatcattgttttggcc 2953
Db   3504 ATTTAACTAAAACTTGAAGAACTAAAACAACAATGCAAACCTTTCAGCATTGTTTGGCC 3563

Qy   2954 aaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaatta 3013
Db   3564 AAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGGCACCAACTAATTA 3623

Qy   3014 tgcaagcatgaattttcacctgagagtgaaaaaagaaaactctaccatggcttgaagtt 3073
Db   3624 -GCAAGCATGAATTTTTCACCCAAGAGTGAAAAAAGGAAAATCTACCATGGCTTGAAGTT 3682

Qy   3074 acaggagcagaactcctgactaccattctatgactgatgaagagactaata-tctaaaac 3132
Db   3683 A-AAGAGCAGAACTCCTGACTACCATTCTATGACTGATCAAAAGACTAATAGTTAAAAAC 3741

Qy   3133 ctcagcaggccttgttcacgatatgcag---aaaaagtgctgcagtttagatacctctgg 3189
Db   3742 CTCAGCAGGCCTTGTTCACGATATGCAGAAAAAAAAGTGCTGCAGTTTAGATACCTCT-G 3800

Qy   3190 gaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaatata 3249
Db   3801 GAATTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTACATTTCTAAGAATATA 3860

Qy   3250 tttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagat 3309
```

```
Db    3861 TTTCTTGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGGGCCCAAGAT 3920

Qy    3310 tttttttt------------caataacagaccagcttctttttcctgcagttacaaatg 3356
Db    3921 TTTTTTTTTTTTTTTTTTTTCAAATAACAGACCAGCTTCTTTTTCTTGCAGTTACAGATG 3980

Qy    3357 taatttctttttttttttgttgtcaaacataaggtaccaaat---atgcaataaattgtt 3413
Db    3981 TAATTTC-----CTTTTTGTTGTCAAACATAAGGTACCAAATATGATGCAATAAATTGTT 4035

Qy    3414 ttg 3416
Db    4036 TTG 4038


RESULT   6
AF327722
LOCUS       AF327722                4192 bp    mRNA    linear   PRI 05-MAR-2001
DEFINITION  Homo sapiens putative acetyltransferase mRNA, complete cds.
ACCESSION   AF327722
VERSION     AF327722.1  GI:13195459
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4192)
  AUTHORS   He,Y.G., Tan,D.Y., Lai,J.H., Xie,Y.F. and Qian,W.
  TITLE     Cloning and analysis of a human putative acetyltransferase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4192)
  AUTHORS   He,Y.G., Tan,D.Y., Lai,J.H., Xie,Y.F. and Qian,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Biology Department, Yunnan University,
            North Street of Greenlake, Kunming, Yunnan 650091, China
FEATURES             Location/Qualifiers
     source          1. .4192
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             280. .2880
                     /codon_start=1
                     /product="putative acetyltransferase"
                     /protein_id="AAK15707.1"
                     /db_xref="GI:13195460"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                     LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                     QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                     LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                     DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                     ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                     YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                     HEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVLRQHPFYFKAARIAIEIYL
                     KLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKAQIEEEKKNAEKEKQQRNQ
                     KKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPLKNLVKNKIETHLFAFEIY
                     FRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFNTAVCESKDLSDTVRTVLKQEMNR
                     LFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDPSSQKRAIELATTLDESLTNRN
                     LQTCMEVLEALYDGSLGDCKEAAEIYRANCHKLFPYALAFMPPGYEEDMKITVNGDSS
                     AEAEELANEI"
BASE COUNT     1427 a    710 c    953 g    1102 t
ORIGIN


  Query Match           74.4%;  Score 2543;  DB 9;  Length 4192;
  Best Local Similarity  89.9%;  Pred. No. 0;
  Matches 2993;  Conservative    0;  Mismatches  295;  Indels   43;  Gaps   23;

Qy      27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81
Db     710 GACCTGCGCAGAGAGCATCATGGATTGGTTATGCTATTGCTTACCATTTATTAGAAGATT 769

Qy      82 atgaaatggcagcaaaaatttagaagagtttaggaaaacacagcagacatctcctgata 141
Db     770 ATGAAATGGCAGCAAAGATTTTAGAAGAATTTAGGAAAACACAACAGACATCCCCTGACA 829

Qy     142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201
Db     830 AGGTGGATTATGAATATAGTGAACTACTCTTATATCAGAATCAAGTTCTTCGGGAAGCAG 889

Qy     202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261
Db     890 GTCTCTATAGAGAAGCTTTGGAACATCTTTGTACCTATGAAAAGCAGATTTGTGATAAAC 949

Qy     262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321
Db     950 TTGCTGTAGAAGAAACCAAAGGGGAACTTCTGTTGCAACTATGTCGTTTGGAAGATGCTG 1009

Qy     322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381
Db    1010 CAGATGTTTATAGAGGATTGCAAGAGAGAAATCCTGAAAACTGGGCCTATTACAAAGGCT 1069

Qy     382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441
Db    1070 TGGAAAAAGCACTCAAGCCAGCTAATATGTTAGAACGGCTAAAAATTTATGAGGAAGCCT 1129

Qy     442 ggactaaatacccccaggggactcgtgccaagaaagctgcccttaaacttttttatctggag 501
Db    1130 GGACTAAATATCCCAGGGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTG 1189

Qy     502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561
Db    1190 AGAAGTTTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGCCCAC 1249

Qy     562 ctgtcttcaataccttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621
Db    1250 CAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGTGGCAATCATAGAAG 1309

Qy     622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681
Db    1310 AGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATG 1369

Qy     682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741
Db    1370 ATGGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATT 1429

Qy     742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801
Db    1430 ATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTA 1489

Qy     802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861
Db    1490 CACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATA 1549

Qy     862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921
Db    1550 TTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTA 1609

Qy     922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981
Db    1610 TCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAA 1669

Qy     982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041
Db    1670 TGTGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGT 1729

Qy    1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101
Db    1730 GCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAG 1789

Qy    1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
Db    1790 CACTTAAGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTG 1849

Qy    1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
```


Db 1850 ACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTTATTAAAAC 1909
Qy 1222 tagaagatgtacttcgacagcatccatttacttcaaagcagcgagaattgctattgaga 1281
Db 1910 TAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGA 1969
Qy 1282 tctatttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgata 1341
Db 1970 TCTATTTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGCTGATA 2029
Qy 1342 cagcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaa 1401
Db 2030 CAGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAA 2089
Qy 1402 agaaagcccagattgaagaagagaaaaaaaatgccgaaaaagaaaagccgcaacggaatc 1461
Db 2090 AGAAAGCCCAGATAGAAGAAGAGAAAAAAAATGCAGAAAAAGAAAAGCAGCAGAGAAATC 2149
Qy 1462 cgaaaaagaaaaaggatgatgatgacgaagaaattggaggccccaaagaagagcttatcc 1521
Db 2150 AGAAAAAGAAGAAGGATGATGATGATGAGGAGATAGGAGGTCCAAAAGAAGAACTTATTC 2209
Qy 1522 ctgagaaactggccaaggttgaaactccattggaagaagctattaagtttttaacaccat 1581
Db 2210 CAGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAGCTATTAAATTTTTAACACCGT 2269
Qy 1582 tgaagaacttggtgaagaacaagatagaaactcatctttttgcctttgagatctacttta 1641
Db 2270 TGAAGAACTTGGTGAAGAACAAGATAGAGACTCATCTTTTTGCCTTTGAGATTTACTTTA 2329
Qy 1642 ggaaagaaaagtttcttttgatgctacaatcagtaaagcgggcatttgctattgattcta 1701
Db 2330 GGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGAGGGCATTTGCTATTGATTCTA 2389
Qy 1702 gtcatccctggcttcatgagtgcatgattcgactctttcattct---gtgtgtgaaagta 1758
Db 2390 GTCATCCCTGGCTTCATGAGTGTATGATTCGTCTCTTTAATACTGCAGTGTGTGAAAGTA 2449
Qy 1759 aagacttacccgaaacagttagaacagtattaaaacaagaaatgaatcgtctttttggag 1818
Db 2450 AAGATTTATCTGATACAGTTAGAACAGTATTAAAACAAGAAATGAATCGTCTTTTTGGAG 2509
Qy 1819 caacaaatccaaagaatttttaatgaaacctttctgaaaaggaattctgattcattgccac 1878
Db 2510 CAACGAATCCAAAGAATTTTAATGAAACTTTTCTGAAAAGGAATTCTGATTCATTGCCAC 2569
Qy 1879 atagattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaa 1938
Db 2570 ACAGATTATCAGCTGCCAAAATGGTATATTACTTAGATCCTTCTAGTCAGAAGCGAGCTA 2629
Qy 1939 tagagctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatgg 1998
Db 2630 TAGAGTTGGCAACAACACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGG 2689
Qy 1999 aagtgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcct 2058
Db 2690 AGGTATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAAATTT 2749
Qy 2059 acagagcaagttgtcataagctttcccttatgctttggctttcatgcctcctggatacg 2118
Db 2750 ATAGAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCCTCCTGGATATG 2809
Qy 2119 aagaggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggcca 2178
Db 2810 AAGAGGATATGAAGATCACAGTTAATGGAGATAGTTCTGCAGAAGCTGAAGAACTGGCCA 2869
Qy 2179 atgaaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtat 2238
Db 2870 ATGAAATTTGAACATCACTAAACAAGCAAATGGAATGACTTTGGACCATATCTAGTATAT 2929
Qy 2239 aatatttttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaat 2298
Db 2930 AATATTTTTGTCACGCACCTGCTGCATTGCTCTAACTTACACAGAATGAGAGGAGTAAAT 2989
Qy 2299 gttcttgccttcaaata--gtcttacgtttttatcctgctg-aaaactatatataaaat 2355
Db 2990 GTTCTTGCCTTCAAATAGTGTTTTACGTTTTTTATCCTGCTGAAAAAGTATATATAAAAT 3049
Qy 2356 atctaacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgag 2415
Db 3050 ATCTAACATTACAGGATAGAGGTTCAGTTTCTTAAAAAATT-AAAGCTGCTAAAATTGAG 3108
Qy 2416 gggtttaaaagaaaaaaaaatccgtatcctattcctaccttccctccccatgtttttaac 2475
Db 3109 TGGTT------AAAAAAGATACCTTATCCTATTCCTCCCCACCCACCCATGTTTTTAAAC 3162
Qy 2476 taatttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatat 2535
Db 3163 TAATTTATATAAAATCTGGAGGCTGTTACAGCTAACAAAGCAGGTGTGTGGCAGAAATAT 3222
Qy 2536 tactttaaatttgtcttgtgagattttgctatatctcagacagcataaataaatgctgtt 2595
Db 3223 TACTTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC----ATAAATGCTGGT 3278
Qy 2596 ttagcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgat 2655
Db 3279 TTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTG-TGGGGCTTTAGCATCTGACTGAT 3337
Qy 2656 tctgttacggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacaga 2715
Db 3338 T-TGTTACGGGGTTGAT-TCTGACCATAGGAAGTATGCAATGTGAATCACTATTTACAGA 3395
Qy 2716 gccgtctacaacacatgcttgacgttgtagagactgggacacatagctaccaagc-ggat 2774
Db 3396 GAAACCTACAACAGATGCTTGATGTTGTAGAAACTGGGACATATAGATACCAAGCAAAAT 3455
Qy 2775 taagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatg 2834
Db 3456 TATAAGAAACCTATAAGGTGTTCAATACG---CTTGTGTTTCCAAAATTCACTGTACATG 3512
Qy 2835 atcagtttggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctca 2894
Db 3513 ATCAG-TTGGTGTTCTTGTACCACAGTTTTTAACTGAAGGAACCAGTTGTAACAGTCTCA 3571
Qy 2895 a-tttaactaaaacttgaagaactaaaataacaatgcaaaccttt-atcattgttttggc 2952
Db 3572 ATTTTAACTAAAACTTGAAGAACTAAAACAACAATGCAAACCTTTCAGCATTGTTGTGGC 3631
Qy 2953 caaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaatt 3012
Db 3632 CAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGGCACCAACTAATT 3691
Qy 3013 atgcaagcatgaatttttcacctgagagtgaa-aaaagaaaactctaccatggcttgaag 3071
Db 3692 A-GCAAGCATGAATTTTTCACCCAAGAGTGAACAAAAGGAAAATCTACCATGGCTTGAAG 3750
Qy 3072 ttacaggagcagaactcctgactaccattctatgactgatgaagagactaata-tctaaa 3130
Db 3751 TTA-AAGAGCAGAACTCCTGACTACCATTCTATGACTGATCAAAAGACTAATAGTTAAAA 3809
Qy 3131 acctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttagatacctctggg 3190
Db 3810 ACCTCAGCAGGC--TGTTCACGATATGCAGAAAAAAAAGTGCTGCAGTTTAGATACTCTG 3867
Qy 3191 aacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaatatat 3250
Db 3868 GAATTTTTCACAGTGT--CCAGGTTTGTATACTTGAAGGCCTAC-TTTCTAAGAATAATT 3924
Qy 3251 ttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagatt 3310
Db 3925 TCGTGCGCAGGTGG--TCAGGGAAGCCCAAGACTTGGAAGTTTAAAGGGCGCAAAATTTG 3982
Qy 3311 ttttttttcaataacagaccagcttctttttc 3341
Db 3983 TGTTTTGTTTCTACAAAACGACAGGCTTTTC 4013

```
RESULT   7
AX285247
LOCUS       AX285247               1779 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 6 from Patent WO0179506.
ACCESSION   AX285247
VERSION     AX285247.1  GI:17045931
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Treatment of ocular neovascularization and related diseases
  JOURNAL   Patent: WO 0179506-A 6 25-OCT-2001;
            Children's Hospital Research Foundation (US)
FEATURES             Location/Qualifiers
     source          1. .1779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      627 a    319 c    385 g    448 t
ORIGIN
```

Query Match 52.0%; Score 1779; DB 6; Length 1779;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1779; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    408 atgttagaacggctaaaaatatatgaggaagcctggactaaataccccaggggactcgtg 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTG 60

Qy    468 ccaagaaagctgcccttaaacttttatctggagagaagtttaaggagtgtttggatagg 527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGG 120

Qy    528 ttcctaaggatgaatttcagcaagggctgtccacctgtcttcaatacctttgaggtcttta 587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTCCTAAGGATGAATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTA 180

Qy    588 tacagagataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttct 647
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACAGAGATAAAGAGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCT 240

Qy    648 ctaaaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccaca 707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTAAAAAGTTGTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 300

Qy    708 ttactttgggtccagtactatttggcacagcattatgataaaattggtcagccatccatt 767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATT 360

Qy    768 gctctggaatacataaatactgcaattgaaagtacaccaacattgatagaactctttctt 827
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTT 420

Qy    828 gtaaaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggat 887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTAAAAGCTAAAATCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGAT 480

Qy    888 gaagcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatg 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAAGCCCAGGCCCTGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATG 540

Qy    948 ttaaaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaagga 1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTAAAAGCCAACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 600

Qy   1008 acttcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgct 1067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCT 660

Qy   1068 caggcatacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgag 1127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAG 720

Qy   1128 agacattttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaag 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGACATTTTATAGAAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAG 780

Qy   1188 atcacccttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatcca 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATCACCCTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCA 840

Qy   1248 ttttacttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaaccct 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTTTACTTCAAAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 900

Qy   1308 ctgacagatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagcta 1367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTA 960

Qy   1368 aagaaactgcgtaataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaa 1427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAA 1020

Qy   1428 aaaaatgccgaaaaagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgac 1487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAAAATGCCGAAAAAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGAC 1080

Qy   1488 gaagaaattggaggccccaaagaagagcttatccctgagaaactggccaaggttgaaact 1547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAAGAAATTGGAGGCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACT 1140

Qy   1548 ccattggaagaagctattaagtttttaacaccattgaagaacttggtgaagaacaagata 1607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCATTGGAAGAAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1200

Qy   1608 gaaactcatctttttgcctttgagatctactttaggaaagaaaagtttctttttgatgcta 1667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTA 1260

Qy   1668 caatcagtaaagcgggcatttgctattgattctagtcatccctggcttcatgagtgcatg 1727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATG 1320

Qy   1728 attcgactctttcattctgtgtgtgaaagtaaagacttacccgaaacagttagaacagta 1787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATTCGACTCTTTCATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTA 1380

Qy   1788 ttaaaacaagaaatgaatcgtctttttggagcaacaaatccaaagaattttaatgaaacc 1847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTAAAACAAGAAATGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACC 1440

Qy   1848 tttctgaaaaggaattctgattcattgccacatagattatcagctgccaaaatggtatat 1907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTTCTGAAAAGGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1500

Qy   1908 tatttagattcttctagtcaaaaacgagcaatagagctggcgacaacacttgatggatcc 1967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCC 1560

Qy   1968 ctcaccaacagaaaccttcagacttgcatggaagtgttggaagccttgtgtgatggtagc 2027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGGAAGCCTTGTGTGATGGTAGC 1620

Qy   2028 ctacgagactgtaaagaagctgccgaagcctacagagcaagttgtcataagcttttccct 2087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTACGAGACTGTAAAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCT 1680

Qy   2088 tatgctttggctttcatgcctcctggatacgaagaggatatgaagatcacagtgaacgga 2147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TATGCTTTGGCTTTCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGA 1740
```

```
Qy    2148 gatagttctgcagaaacggaagaactggccaatgaaatc 2186
           |||||||||||||||||||||||||||||||||||||||
Db    1741 GATAGTTCTGCAGAAACGGAAGAACTGGCCAATGAAATC 1779
```

RESULT 8
AF247679
LOCUS AF247679 3324 bp mRNA linear VRT 02-AUG-2000
DEFINITION Xenopus laevis putative N-terminal acetyltransferase mRNA, complete cds.
ACCESSION AF247679
VERSION AF247679.1 GI:9651962
KEYWORDS .
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 3324)
AUTHORS Choi,S.-C., Kim,J. and Han,J.-K.
TITLE Expression of N-terminal acetyltransferase in Xenopus laevis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3324)
AUTHORS Choi,S.-C., Kim,J. and Han,J.-K.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Life Science, Pohang University of Science and Technology, Pohang 790-784, South Korea
FEATURES Location/Qualifiers
source 1. .3324
/organism="Xenopus laevis"
/db_xref="taxon:8355"
CDS 105. .2645
/codon_start=1
/product="putative N-terminal acetyltransferase"
/protein_id="AAF91333.1"
/db_xref="GI:9651963"
/translation="MPSVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
HGETLAMKGLTLNCLGKKEDAYDLVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
YRNALKWDKDNLQILRDLSLLQLRPAQRASWIGYAIAYHLLEDYEMAAKILEEFRKTQ
QTSPDKVDYEYSELLLYQNQVLREAELYKEALDHLCTYEKQICDKLAVEETKGDLLNR
LGRLDEAADVYRALQDRCPENWAYYKGLEKALKPESMEERLKIYEDSWTKYPRGLVSR
RLPLNFLSGLKFRECLDKYLRMNFSKGCPPVFNTLRSLYSDKEKVEIIEDLVVGYETS
LKSCRLFNMNDDGKEEPPTTLLWVQYYLAQHYDKIGQQSIALEYINAAIDSTPTLIEL
FLVKAKIYKHSGNIKEAVRWMDEAQALDTADRFINSKCAKYMLKANLIKEAEEMCSKF
TREGTSAVENLNEMQCMWFQTECAQAYKSMNKYGEALKKCHEIERHFVEITDDQFDFH
TYCMRKITLRSYADLLKLEDVLRQHPFYFKAARIAIEIYLKLHDNPLTDENKEHEADT
ANMSDKELKKLRNKQRRAQKKAQLEEEKKNAEKEKQQRNQKKKKEDDDEEIGGPKEEL
IPEKLAKVENPLEEAIKFLTPLKNLVKNKIETHLYAFEIYFRKDKFLLMLQSVKRAYA
IDPNHPWLHQCLIRFFCAVSESKELNESVRTVLKQEMCRLFGETSPANFNNSFLKENI
NSIPHRLAAAKMMYYLDHSSQKRSVELGTSLDESLCNGSLQTCTDVLEALRDGSLGDK
EAAECYRVSCHKLYPYALAFMPPGYEEDMSITMNGDSSAETEELANEM"
BASE COUNT 1111 a 599 c 756 g 858 t
ORIGIN

Query Match 43.0%; Score 1469.4; DB 5; Length 3324;
Best Local Similarity 75.9%; Pred. No. 6.3e-284;
Matches 1936; Conservative 0; Mismatches 571; Indels 44; Gaps 8;

```
Qy      30 ctgcagagtgagcatcatggattggttatgc-----tttaccatttattagaagactatg 84
           ||||| || ||||||| |||||||| |||||      |||||||||  | ||||||||||
Db     484 CTGCACAGAGAGCATCTTGGATTGGATATGCAATTGCTTACCATTTGCTTGAAGACTATG 543

Qy      85 aaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgataaaag 144
           ||||||| ||||||||  | || |||||| | ||||| |||||||| ||||| || ||||
Db     544 AAATGGCTGCAAAAATACTTGAGGAGTTTCGTAAAACCCAGCAGACTTCTCCAGACAAAG 603

Qy     145 tggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcaggtc 204
           |||| |||||||| ||||| || ||  | ||||| ||||||||||  | || ||||  |
Db     604 TGGACTATGAATACAGTGAGCTGCTGCTTTATCAAAATCAAGTTCTCAGAGAGGCAGAGC 663

Qy     205 tttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaacttg 264
           | |||| ||| || ||||| |||||||| |||||||||||||| || |||||||| || |
```

```
Db     664 TCTATAAAGAGGCTCTGGATCATCTTTGCACCTATGAAAAGCAAATCTGTGATAAGCTAG 723

Qy     265 ctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctgctg 324
           | |||||||||||| ||||| ||  | |||  |  ||  | ||  | || || || || |
Db     724 CCGTTGAAGAAACAAAAGGAGATTTGCTGAATCGTTTAGGCCGACTTGATGAGGCAGCGG 783

Qy     325 acgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggcttag 384
           | || || || |  || || || ||    |||||||| ||||| |||||||| || ||||
Db     784 ATGTCTACAGGGCGTTGCAGGACAGATGCCCTGAAAACTGGGCATATTACAAGGGTTTAG 843

Qy     385 aaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcctgga 444
           |||| ||||| || ||||  | ||||  |||| | || ||||| ||||| ||   |||||
Db     844 AAAAGGCACTTAAACCAGAAAGTATGGAAGAAAGACTGAAAATTTATGAAGACAGCTGGA 903

Qy     445 ctaaataccccaggggactcgtgccaagaaagctgcccttaaacttttatctggagaga 504
           | || || |||||||||| | ||  ||||||  ||||| |||||||| ||||| ||    |
Db     904 CAAAGTATCCCAGGGGATTAGTTTCAAGAAGACTGCCTTTAAACTTCTTATCAGGTCTAA 963

Qy     505 agtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccacctg 564
           | ||||  || ||| ||||||  |  | ||||||||||| ||||| || || || || |
Db     964 AATTTAGAGAATGTCTGGATAAATATTTGAGGATGAATTTTAGCAAAGGATGCCCTCCAG 1023

Qy     565 tcttcaatacctttgaggtctttatacagagataaagagaaggtggcaatcgtagaagaac 624
           | || ||||| ||||| ||  | ||||| |||||||||||||| |  ||  | |||||
Db    1024 TATTTAATACATTGAGATCACTCTACAGCGATAAAGAGAAGGTAGAGATTATTGAAGATT 1083

Qy     625 tagtagtttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatgatg 684
           |||| |||||||||||||| ||  | ||||| || ||  ||||||||  ||||||||||
Db    1084 TAGTTGTTGGTTATGAAACATCCTTGAAAAGCTGCCGTTTATTTAACATGAATGATGATG 1143

Qy     685 gaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcattatg 744
           | || || ||||| ||||| || ||||| ||||| |||||||| ||||||| |||||||
Db    1144 GGAAAGAAGAACCACCAACAACTTTACTGTGGGTGCAGTACTACTTGGCACAACATTATG 1203

Qy     745 ataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagtacac 804
           | |||||||| |||| ||| |||||||||||||| || ||| |||| ||||| || ||||
Db    1204 ACAAAATTGGCCAGCAATCAATTGCTCTGGAATATATTAATGCTGCGATTGACAGCACAC 1263

Qy     805 caacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaatatta 864
           ||||  | || ||||| || ||||| |||||||||||| || |||||| |||||||||| |
Db    1264 CAACTCTTATTGAACTTTTCCTTGTCAAAGCTAAAATTTACAAGCATTCTGGGAATATAA 1323

Qy     865 aagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagatttatta 924
           ||||||||| | | |||||||||||||| ||||| ||||| ||||| ||| |||||||| |
Db    1324 AAGAAGCTGTCCGATGGATGGATGAAGCGCAGGCACTGGATACAGCTGACCGATTTATCA 1383

Qy     925 attccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaatgt 984
           |||| || |||||||||||| |||||||||| |||||||||||||| ||||||||||||
Db    1384 ATTCTAAATGTGCAAAATATATGTTAAAAGCAAACCTGATTAAAGAAGCTGAAGAAATGT 1443

Qy     985 gttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagtgta 1044
           | || || ||||| |||||||||||| |||||| || || || |||||||| |||||||| |
Db    1444 GCTCGAAATTTACAAGGGAAGGAACATCAGCAGTGGAAAATCTGAATGAGATGCAGTGCA 1503

Qy    1045 tgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaagcac 1104
           |||||||||||||||| ||||| || || ||||  | ||||| |||| ||||||||||
Db    1504 TGTGGTTCCAGACAGAATGTGCACAAGCTTACAAGTCCATGAATAAATATGGTGAAGCAC 1563

Qy    1105 ttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttgact 1164
           |||| ||||| |||||||||| |||||||| ||||||| || ||||||||||||| |
Db    1564 TTAAAAAATGCCATGAAATTGAAAGACATTTTGTAGAAATTACAGATGACCAGTTTGATT 1623

Qy    1165 ttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaactag 1224
           | || || |||||||||| || || || || |||||||| ||||||||||||| | |
Db    1624 TCCACACTTACTGTATGAGAAAAATTACACTCAGGTCATATGCGGACTTATTAAAATTGG 1683

Qy    1225 aagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgagatct 1284
           | ||||||||  | |||||||||||||||||||| || ||  | ||||| || || || |
Db    1684 AGGATGTACTAAGGCAGCATCCATTTTACTTCAAGGCTGCACGGATTGCAATAGAAATTT 1743

Qy    1285 atttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgatacag 1344
           || |||||||||||| || ||||| |||||||| || ||||| || |||| || ||||
Db    1744 ATCTGAAGCTTCATGATAATCCTCTAACAGATGAAAATAAAGAGCATGAGGCAGACACAG 1803
```

```
Qy   1345 caaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaaaga 1404
          | || ||||||||||||||||| |||||||| ||||| |||||| | || || || ||||
Db   1804 CCAATATGTCTGACAAAGAGCTGAAGAAACTACGTAACAAACAACGGAGGGCGCAGAAGA 1863

Qy   1405 aagcccagattgaagaagagaaaaaaaatgccgaaaaagaaaagccgcaacggaatccga 1464
          |||| ||| | ||||| |||||||| || || |||||||||||||  ||  | || |  |
Db   1864 AAGCACAGCTGGAAGAGGAGAAAAAGAACGCGGAAAAAGAAAAGCAACAGAGAAACCAAA 1923

Qy   1465 aaaagaaaaaggatgatgatgacgaagaaattggaggccccaaagaagagcttatccctg 1524
          | |||||||| || |||||||| ||||| |||||||| || || ||||||||||| || |
Db   1924 AGAAGAAAAAAGAAGATGATGATGAAGAGATTGGAGGACCGAAGGAAGAGCTTATACCAG 1983

Qy   1525 agaaactggccaaggttgaaactccattggaagaagctattaagtttttaacaccattga 1584
          | ||  | || |||||||||| ||| |||||||||||||||||| ||||||||||  | |
Db   1984 AAAAGTTAGCAAAGGTTGAAAATCCTTTGGAAGAAGCTATTAAATTTTTAACACCTCTAA 2043

Qy   1585 agaacttggtgaagaacaagatagaaactcatctttttgcctttgagatctactttagga 1644
          |||||||||||||||||| ||||| || ||||||| |||||||| || ||||| ||||
Db   2044 AGAACTTGGTGAAGAACAAAATAGAGACACATCTTTATGCCTTTGAAATTTACTTCAGGA 2103

Qy   1645 aagaaaagtttcttttgatgctacaatcagtaaagcgggcatttgctattgattctagtc 1704
          |||| ||||||||| ||||| | || || || || || |||| || || ||| ||| |
Db   2104 AAGATAAGTTTCTTCTGATGTTGCAGTCTGTGAAACGAGCATACGCAATAGATCCTAACC 2163

Qy   1705 atccctggcttcatgagtgcatgattcgactctttcattctgtgtgtgaaagtaaagact 1764
          | || |||||||||| ||||  | || ||| |||||  | | || | ||||||||||||
Db   2164 ACCCGTGGCTTCATCAGTGTCTTATCCGATTCTTTTGTGCAGTTTCTGAAAGTAAAGAAC 2223

Qy   1765 tacccgaaacagttagaacagtattaaaacaagaaatgaatcgtctttttggagcaacaa 1824
          ||  |||| | || |||||||| ||||| |||||||||     | ||||||||||| ||| |
Db   2224 TAAACGAATCGGTAAGAACAGTGTTAAAGCAAGAAATGTGCAGACTTTTTGGAGAAACCA 2283

Qy   1825 atccaaagaattttaatgaaaccttttctgaaaaggaattctgattcattgccacatagat 1884
            ||    ||||||||| |  |||| ||||||   ||     | ||  | |||||  |||
Db   2284 GCCCTGCAAATTTTAATAATTCCTTCCTGAAAGAAAACATCAACTCCATACCACACCGAT 2343

Qy   1885 tatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaatagagc 1944
          || ||||||||||||||| | || ||| | ||   ||| ||||||||||| || ||||||
Db   2344 TAGCAGCTGCCAAAATGATGTACTATCTGGACCATTCAAGTCAAAAACGATCAGTAGAGC 2403

Qy   1945 tggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatggaagtgt 2004
          ||| |||  || | |||| ||| |||   ||| | |  || ||||| ||||  || ||
Db   2404 TGGGGACCTCATTGGATGAATCGCTCTGTAACGGTAGTCTCCAGACGTGCACCGATGTTC 2463

Qy   2005 tggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcctacagag 2064
          |||||||||||  | || || ||  || |||||   ||||| ||||| |||   ||  |||
Db   2464 TGGAAGCCTTGCGAGACGGCAGTTTAGGAGAC---AAAGAGGCTGCTGAATGTTATCGAG 2520

Qy   2065 caagttgtcataagcttttcccttatgctttggctttcatgcctcctggatacgaagagg 2124
            ||||| || ||||| | |||||| ||||| |||||||||||||||| ||||||||||
Db   2521 TCAGTTGCCACAAGCTCTACCCTTACGCTTTAGCTTTCATGCCTCCTGGCTACGAAGAGG 2580

Qy   2125 atatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggccaatgaaa 2184
          | ||||  |||||| |||| |||||||| || || ||||| ||||||||||  |||||||
Db   2581 ACATGAGTATCACAATGAATGGAGATAGCTCCGCCGAAACAGAAGAACTGGCTAATGAAA 2640

Qy   2185 tctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtataatatt 2244
          | |||| | |    ||  | ||        ||| |||||||||||  | |  | |||||||
Db   2641 TGTGAATAGCTGATAAACAACA--------GACATTGGACCATATCAACTATCTAATATT 2692

Qy   2245 tttgtcacgcacctgctgcattgctct----tacttacacagaatgagaggagtaaatg- 2299
          |||||||||||||||| || |||||        |    ||||||| |||||||||  |
Db   2693 TTTGTCACGCACCTGCTGTATCGCTCTGTAACTTATGTGGAGAATGAAAGGAGTAAAAGA 2752

Qy   2300 -ttcttgccttcaaata--gtcttacgtttttttatcctgctgaaaactatatataaaata 2356
           || ||||||| |||||  || |||| ||||||||         |  ||  | || |  |
Db   2753 ATTGTTGCCTTTAAATATTGTTTTACTTTTTTTATATCCTGCTGAGGTAGTTCTATATAA 2812

Qy   2357 tctaacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgagg 2416
          | ||  |   |     ||| |  |||||||||||   |   |||||||||  |  ||
Db   2813 TGTATTACAGGATATGGGAGGATTGGTTTCTTAAAAATAAATTTGCTGCTAAACATTTGG 2872

Qy   2417 ggtttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaact 2476
          |      ||  | ||| ||| ||||    ||||                 |||||| || |
Db   2873 TG------AGGGAGAAACATCTGTATGGAGTTCCC-------------TGTTTTAAATT 2912

Qy   2477 aatttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatatt 2536
          |||||||| | |  |  |||   | ||||  | || ||||||||||||||| |||||||
Db   2913 AATTTATATGAGAGATAAAGGTGGTTACAGTCACCAGAGCAGGTGTGTGGCACAAATATT 2972

Qy   2537 actttaaatttgtcttgtgagattttgctat 2567
          |||    ||    ||  ||||  || || |
Db   2973 ACTAACGTTTGTCTTTTAGAGAGATTACTTT 3003


RESULT   9
AX285295/c
LOCUS       AX285295               1413 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 3 from Patent WO0179505.
ACCESSION   AX285295
VERSION     AX285295.1  GI:17045976
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 3 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .1413
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      359 a    314 c    244 g    496 t
ORIGIN


  Query Match           41.3%;  Score 1413;  DB 6;  Length 1413;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-272;
  Matches 1413;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1413 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 1354

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaatttagaagagtttaggaaaa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1353 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 1294

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1293 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 1234

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1233 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 1174

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaaggggaacttctgttgcagt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1173 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 1114

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1113 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 1054

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1053 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 994

Qy     421 taaaaatatatgaggaagcctggactaaatacccccaggggactcgtgccaagaaagctgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     993 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 934
```

```
Qy    481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 874

Qy    541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 814

Qy    601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 754

Qy    661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 694

Qy    721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaatacа 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 634

Qy    781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 574

Qy    841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 514

Qy    901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 454

Qy    961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 394

Qy   1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 334

Qy   1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 274

Qy   1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 214

Qy   1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccattttacttcaaag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 154

Qy   1261 cagcgagaattgctattgagatctatttgaagcttcatgacaacccctctgacagatgaga 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 94

Qy   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 34

Qy   1381 ataaacaaagaagagctcaaaagaaagcccaga 1413
          ||||||||||||||||||||||||||||||||
Db     33 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGA 1

RESULT  10
AK023387
LOCUS       AK023387                1802 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ13325 fis, clone OVARC1001762, weakly similar
            to N-TERMINAL ACETYLTRANSFERASE 1 (EC 2.3.1.88).
ACCESSION   AK023387
VERSION     AK023387.1  GI:10435302
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_lib:OVARC1
            clone:OVARC1001762.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1802)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="OVARC1001762"
                     /clone_lib="OVARC1"
                     /tissue_type="ovary, tumor tissue"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      644 a    298 c    395 g    465 t
ORIGIN

  Query Match           35.4%;  Score 1210.2;  DB 9;  Length 1802;
  Best Local Similarity  92.3%;  Pred. No. 4.7e-232;
  Matches 1287;  Conservative    0;  Mismatches  103;  Indels    5;  Gaps    1;

Qy     27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81
          || ||||  || ||||||||||||||||||||||     ||||||||||||||||||| |
Db    408 GACCTGCGCAGAGAGCATCATGGATTGGTTATGCTATTGCTTACCATTTATTAGAAGATT 467

Qy     82 atgaaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgata 141
          |||||||||||||| |||||||||||| |||||||||||| |||||||| ||||| |
Db    468 ATGAAATGGCAGCAAAGATTTTAGAAGAATTTAGGAAAACACAACAGACATCCCCTGACA 527

Qy    142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201
          | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    528 AGGTGGATTATGAATATAGTGAACTACTCTTATATCAGAATCAAGTTCTTCGGGAAGCAG 587

Qy    202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261
          |||| ||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db    588 GTCTCTATAGAGAAGCTTTGGAACATCTTTGTACCTATGAAAAGCAGATTTGTGATAAAC 647

Qy    262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321
          ||||||| |||||||||||||||||||||||| | ||||||||||||||||||||||||
Db    648 TTGCTGTAGAAGAAACCAAAGGGGAACTTCTGTTGCAACTATGTCGTTTGGAAGATGCTG 707

Qy    322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381
          | || |||||||||||||| |||||||| |||||||||| ||||||||||||||||||||
Db    708 CAGATGTTTATAGAGGATTGCAAGAGAGAAATCCTGAAAACTGGGCCTATTACAAAGGCT 767

Qy    382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441
          | |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db    768 TGGAAAAAGCACTCAAGCCAGCTAATATGTTAGAACGGCTAAAAATTTATGAGGAAGCCT 827
```

```
Qy     442 ggactaaataccccaggggactcgtgccaagaaagctgccctaaacttttatctggag 501
           ||||||||| |||||||||| |||||||||| |||||| |||||||||||||||||| |
Db     828 AGACTAAATATCCCAGGGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTG 887

Qy     502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561
           |||||||||| || |||||||||| ||||||||||||||||||||||||||| || ||||
Db     888 AGAAGTTTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGCCCAC 947

Qy     562 ctgtcttcaataccttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621
           | |||||||||||| || || || ||||||| ||| ||||| |||||||||||| ||||||
Db     948 CAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGTGGCAATCATAGAAG 1007

Qy     622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681
           | ||||||| |||||||||| |||||||||| || ||  |||||||||||||||||||
Db    1008 AGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATG 1067

Qy     682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741
           |||||||||||||||| ||||||||||||||||||||||||||||| |||||||| ||||
Db    1068 ATGGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATT 1127

Qy     742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801
           |||| |||||||||||||||| |||||| |||| |||||||||||||| ||||||||||
Db    1128 ATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTA 1187

Qy     802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861
           |||| ||||| |||||||||||||| || ||||||||||||||||||||||||| ||||
Db    1188 CACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATA 1247

Qy     862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921
           |||||||||||||| |||||||||||||| |||||||| ||||||||||||||||||||
Db    1248 TTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTA 1307

Qy     922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981
           | || ||||| |||||||||||||| ||||||||||||||||||||| ||||||||||
Db    1308 TCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAA 1367

Qy     982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041
           |||| || |||||||| |||||||||| |||||||||||||  ||||||||||||||||
Db    1368 TGTGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGT 1427

Qy    1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101
           | |||||||||||| ||||| ||||| ||||| || |||||||||| ||||||||||||
Db    1428 GCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAG 1487

Qy    1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
           |||||||||||||||| || |||||||||||||||||||||||| ||||||||||||||
Db    1488 CACTTAAGAAATGTCACGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTG 1547

Qy    1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1548 ACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTTATTAAAAC 1607

Qy    1222 tagaagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgaga 1281
           |||||||||||||||||||||||||||||||||||| ||||| |||||||||| ||||
Db    1608 TAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGA 1667

Qy    1282 tctatttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgata 1341
           |||||||||||||||||||||||||| || |||||||||| |||||||||| |||||||
Db    1668 TCTATTTGAAGCTTCATGACAACCCCCTTACAGATGAGAATAAAGAACACGAAGCTGATA 1727

Qy    1342 cagcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaa 1401
           |||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    1728 CAGCAAACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAA 1787

Qy    1402 agaaagcccagattg 1416
           ||||||||||||| |
Db    1788 AGAAAGCCCAGATAG 1802
```

```
RESULT  11
AC094440/c
LOCUS       AC094440              101276 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-4F1, *** SEQUENCING IN PROGRESS ***,
            51 unordered pieces.
ACCESSION   AC094440
VERSION     AC094440.2  GI:17941168
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 101276)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101276)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:15624274.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAPL
                Center clone name: CH230-4F1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 81999 bases at least Q40
                Consensus quality: 88613 bases at least Q30
                Consensus quality: 93921 bases at least Q20
                Estimated insert size: 77412; sum-of-contigs estimation
```

```
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     4304: contig of 4304 bp in length
*     4305     4404: gap of unknown length
*     4405     7807: contig of 3403 bp in length
*     7808     7907: gap of unknown length
*     7908    11854: contig of 3947 bp in length
*    11855    11954: gap of unknown length
*    11955    15183: contig of 3229 bp in length
*    15184    15283: gap of unknown length
*    15284    17669: contig of 2386 bp in length
*    17670    17769: gap of unknown length
*    17770    19299: contig of 1530 bp in length
*    19300    19399: gap of unknown length
*    19400    22678: contig of 3279 bp in length
*    22679    22778: gap of unknown length
*    22779    26720: contig of 3942 bp in length
*    26721    26820: gap of unknown length
*    26821    30932: contig of 4112 bp in length
*    30933    31032: gap of unknown length
*    31033    33143: contig of 2111 bp in length
*    33144    33243: gap of unknown length
*    33244    35654: contig of 2411 bp in length
*    35655    35754: gap of unknown length
*    35755    38153: contig of 2399 bp in length
*    38154    38253: gap of unknown length
*    38254    40769: contig of 2516 bp in length
*    40770    40869: gap of unknown length
*    40870    42868: contig of 1999 bp in length
*    42869    42968: gap of unknown length
*    42969    45108: contig of 2140 bp in length
*    45109    45208: gap of unknown length
*    45209    47511: contig of 2303 bp in length
*    47512    47611: gap of unknown length
*    47612    49283: contig of 1672 bp in length
*    49284    49383: gap of unknown length
*    49384    51138: contig of 1755 bp in length
*    51139    51238: gap of unknown length
*    51239    52352: contig of 1114 bp in length
*    52353    52452: gap of unknown length
*    52453    54071: contig of 1619 bp in length
*    54072    54171: gap of unknown length
*    54172    55553: contig of 1382 bp in length
*    55554    55653: gap of unknown length
*    55654    57107: contig of 1454 bp in length
*    57108    57207: gap of unknown length
*    57208    58972: contig of 1765 bp in length
*    58973    59072: gap of unknown length
*    59073    60489: contig of 1417 bp in length
*    60490    60589: gap of unknown length
*    60590    61766: contig of 1177 bp in length
*    61767    61866: gap of unknown length
*    61867    63712: contig of 1846 bp in length
*    63713    63812: gap of unknown length
*    63813    65491: contig of 1679 bp in length
*    65492    65591: gap of unknown length
*    65592    67146: contig of 1555 bp in length
*    67147    67246: gap of unknown length
*    67247    68788: contig of 1542 bp in length
*    68789    68888: gap of unknown length
*    68889    70062: contig of 1174 bp in length
*    70063    70162: gap of unknown length
*    70163    72007: contig of 1845 bp in length
*    72008    72107: gap of unknown length
*    72108    73211: contig of 1104 bp in length
*    73212    73311: gap of unknown length
*    73312    74467: contig of 1156 bp in length
*    74468    74567: gap of unknown length
*    74568    75716: contig of 1149 bp in length
*    75717    75816: gap of unknown length
*    75817    76865: contig of 1049 bp in length
*    76866    76965: gap of unknown length
*    76966    78104: contig of 1139 bp in length
*    78105    78204: gap of unknown length
*    78205    80369: contig of 2165 bp in length
*    80370    80469: gap of unknown length
*    80470    82200: contig of 1731 bp in length
*    82201    82300: gap of unknown length
*    82301    83833: contig of 1533 bp in length
*    83834    83933: gap of unknown length
*    83934    85135: contig of 1202 bp in length
*    85136    85235: gap of unknown length
*    85236    86701: contig of 1466 bp in length
*    86702    86801: gap of unknown length
*    86802    88323: contig of 1522 bp in length
*    88324    88423: gap of unknown length
*    88424    89873: contig of 1450 bp in length
*    89874    89973: gap of unknown length
*    89974    91605: contig of 1632 bp in length
*    91606    91705: gap of unknown length
*    91706    93024: contig of 1319 bp in length
*    93025    93124: gap of unknown length
*    93125    94257: contig of 1133 bp in length
*    94258    94357: gap of unknown length
*    94358    95684: contig of 1327 bp in length
*    95685    95784: gap of unknown length
*    95785    96826: contig of 1042 bp in length
*    96827    96926: gap of unknown length
*    96927    98126: contig of 1200 bp in length
*    98127    98226: gap of unknown length
*    98227    99641: contig of 1415 bp in length
*    99642    99741: gap of unknown length
*    99742   101276: contig of 1535 bp in length.
FEATURES             Location/Qualifiers
     source          1. .101276
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-4F1"
BASE COUNT    25958 a   21889 c   21793 g   26439 t    5197 others
ORIGIN


  Query Match           34.9%;  Score 1191.8;  DB 2;  Length 101276;
  Best Local Similarity  94.9%;  Pred. No. 1.4e-228;
  Matches 1367;  Conservative    0;  Mismatches   47;  Indels   27;  Gaps   12;

Qy    1983 cttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgtaaa 2042
           ||  ||||  |||||||| |||||||||||||||||||| ||||||| || ||||||||||
Db   40250 CTGTAGACGTGCATGGAGGTGTTGGAAGCCTTGTGTGGTGGTAGCTTAGGAGACTGTAAA 40191

Qy    2043 gaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctttc 2102
           |||||||| ||||| || ||||  ||||||||||||||||||||||||||||| ||||||||
Db   40190 GAAGCTGCTGAAGCGTATAGAGTGAGTTGTCATAAGCTTTTCCCTTATGCTCTGGCTTTC 40131

Qy    2103 atgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcagaa 2162
           |||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||
Db   40130 ATGCCTCCTGGATATGAAGAGGATATGAAGATCACAGTTAACGGAGATAGTTCTGCAGAA 40071

Qy    2163 acggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgactttgg 2222
           |||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db   40070 ACGGAAGAACTGGCCAATGAAATTTGAACATCATTAAACAAGCAAACGGAATGACTTTGG 40011

Qy    2223 accatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacacag 2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40010 ACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTACACAG 39951
```

```
Qy   2283 aatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgctgaaaa 2342
          ||||||||||||||||||||||||||||||     |    ||||||||||||||||||||
Db  39950 AATGAGAGGAGTAAATGTTCTTGCCTTCAATAGTCTTACGTTTTTTTATCCTGCTGAAAA 39891

Qy   2343 ctatatataaaatatc-taacattacaggatataggttcagtttcttaaaaaattaaaag 2401
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  39890 CTATATATAAAATATCTTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTAAAAG 39831

Qy   2402 ctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccctt 2461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39830 CTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTCCCTT 39771

Qy   2462 cccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggtg 2521
          |||||||||||||||||||||||  | ||||||||||||||||||||||||||||||||||
Db  39770 CCCATGTTTTTAACTAATTTATAT--ACTCTGGAGGCTATAACAGCTAACATAGCAGGTG 39713

Qy   2522 tgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagcat 2581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39712 TGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC-- 39653

Qy   2582 aaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggcttt 2641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39654 --ATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTTT 39597

Qy   2642 agcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtgaa 2701
          |||||||| |||||| |||||||||||||| | ||||||||||||||||||||||||||||
Db  39596 AGCATCTGACTGATTTTGTTACGGGGTTGATTATTGACCATAGGAAGTATGCAATGTGAA 39537

Qy   2702 tcactgtgtacagagccgtctacaacacatgcttgacgttgtaga-gactgggacacata 2760
          |||||||||||||||| |||||||| |||||||||||||||||  ||| |||||||||
Db  39536 TCACTGTGTACAGAGCCATCTACAACAGATGCTTGACGTTGTAGATAACTTGGACACATA 39477

Qy   2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
          |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  39476 GCTACCAAGCAGATTAAATGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAA 39417

Qy   2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39416 ATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAG 39357

Qy   2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  39356 TTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAAGCTTTAG 39297

Qy   2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39296 CATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGG 39237

Qy   3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
          |||||||||||| |||||||||| |||||||||||||||||||||||| ||||||||||
Db  39236 CACCAACTAATTA-GCAAGCATGACTTTTTCACCTGAGAGTGAAAAAAGGAAACTCTACC 39178

Qy   3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
          |||||||||||||| | ||||||||||||||||||||||      ||||| ||||||||
Db  39177 ATGGCTTGAAGTTA-AAGAGCAGAACTCCTGACTACCATT------CTGATCAAGAGACT 39125

Qy   3121 aatatct-aaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttag 3179
          |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39124 AATAGCTAAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTTAG 39065

Qy   3180 atacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttc 3239
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39064 ATACCTCT-GGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTC 39006

Qy   3240 taagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaag 3299
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  39005 TAAGAATATATTTCTTGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAG 38946

Qy   3300 ggcccaagattttt-ttttcaataacagaccagcttctttttcctgcagttacaaatgta 3358
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  38945 GGCCCAAGATTTTTCTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGTA 38886

Qy   3359 atttctttttttttttgttgtcaaacataaggtaccaaat---atgcaataaattgtttt 3415
          |||||     ||||||||||||||||||||||||||||||   ||||||||||||||||
Db  38885 ATTTC-----TTTTTTGTTGTCAAACATAAGGTACCAAATATGATGCAATAAATTGTTTT 38831

Qy   3416 g 3416
          |
Db  38830 G 38830
```

RESULT 12
AK001595
LOCUS AK001595 2859 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ10733 fis, clone NT2RP3001392.
ACCESSION AK001595
VERSION AK001595.1 GI:7022944
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA, clone_lib:NT2RP3 clone:NT2RP3001392.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2859)
AUTHORS Isogai,T. and Otsuki,T.
TITLE. Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES Location/Qualifiers
source 1. .2859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RP3001392"
/clone_lib="NT2RP3"
/note="cloning vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."
BASE COUNT 921 a 476 c 547 g 915 t
ORIGIN

Query Match 32.9%; Score 1126; DB 9; Length 2859;
Best Local Similarity 90.2%; Pred. No. 3.1e-215;
Matches 1407; Conservative 0; Mismatches 105; Indels 48; Gaps 17;

```
Qy   1882 gattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaatag 1941
          |||||||||||||||||||||||||||||| |||||| |||||||||| || ||||| ||||
Db      1 GATTATCAGCTGCCAAAATGGTATATTACTTAGATCCTTCTAGTCAGAAGCGAGCTATAG 60

Qy   1942 agctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatggaag 2001
          || |||| |||||||||||||| ||| |||||| |||||||||||| |||||| || ||||| |
Db     61 AGTTGGCAACAACACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGGAGG 120

Qy   2002 tgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcctaca 2061
```

```
Db    121 TATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAAATTTATA 180

Qy   2062 gagcaagttgtcataagcttttcccttatgctttggctttcatgcctcctggatacgaag 2121

Db    181 GAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCCTCCTGGATATGAAG 240

Qy   2122 aggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggccaatg 2181

Db    241 AGGATATGAAGATCACAGTTAATGGAGATAGTTCTGCAGAAGCTGAAGAACTGGCCAATG 300

Qy   2182 aaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtataat 2241

Db    301 AAATTTGAACATCACTAAACAAGCAAATGGAATGACTTTGGACCATATCTAGTATATAAT 360

Qy   2242 atttttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaatgtt 2301

Db    361 ATTTTTGTCACGCACCTGCTGCATTGCTCTAACTTACACAGAATGAGAGGAGTAAATGTT 420

Qy   2302 cttgccttcaaata--gtcttacgttttttatcctgctg-aaaactatatataaaatatc 2358

Db    421 CTTGCCTTCAAATAGTGTTTTACGTTTTTTATCCTGCTGAAAAAGTATATATAAAATATC 480

Qy   2359 taacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgagggg 2418

Db    481 TAACATTACAGGATAGAGGTTCAGTTTCTTAAAAAATT-AAAGCTGCTAAAATTGAGTGG 539

Qy   2419 tttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaactaa 2478

Db    540 TT------AAAAAAGATACCTTATCCTATTCCTCCCCACCCACCCATGTTTTTAAACTAA 593

Qy   2479 tttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatattac 2538

Db    594 TTTATATAAAATCTGGAGGCTGTTACAGCTAACAAAGCAGGTGTGTGGCAGAAATATTAC 653

Qy   2539 tttaaatttgtcttgtgagattttgctatatctcagacagcataaataaatgctgtttta 2598

Db    654 TTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC----ATAAATGCTGTTTTA 709

Qy   2599 gcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgattct 2658

Db    710 GCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTTTAGCATCTGACTGATTTT 769

Qy   2659 gttacggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacagagcc 2718

Db    770 GTTACGGGGTTGAT-TCTGACCATAGGAAGTATGCAATGTGAATCACTATTTACAGAGAA 828

Qy   2719 gtctacaacacatgcttgacgttgtagagactgggacacatagctaccaagc-ggattaa 2777

Db    829 ACCTACAACAGATGCTTGATGTTGTAGAAACTGGGACATATAGATACCAAGCAAAATTAT 888

Qy   2778 gtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatgatc 2837

Db    889 AAGAAACCTATAAGGTGTTCAATACG---CTTGTGTTTCCAAAATTCACTGTACATGATC 945

Qy   2838 agtttggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctcaa-t 2896

Db    946 AGTTTGGTGTTCTTGTACCACAGTTTTTAACTGAAGGAACCAGTTGTAACAGTCTCAATT 1005

Qy   2897 ttaactaaaacttgaagaactaaaataacaatgcaaacctttatcattgttttggccaaa 2956

Db   1006 TTAACTAAAACTTGAAGAACTAAAACAACAATGCAAACCTTTCAGCATTGTTTGGCCAAA 1065

Qy   2957 cttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaattatgc 3016

Db   1066 CTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGGCACCAACTAATTA-GC 1124

Qy   3017 aagcatgaatttttcacctgagagtgaaaaaagaaaactctaccatggcttgaagttaca 3076

Db   1125 AAGCATGAATTTTTCACCCAAGAGTGAAAAAAGGAAAATCTACCATGGCTTGAAGTTA-A 1183

Qy   3077 ggagcagaactcctgactaccattctatgactgatgaagagactaata-tctaaaacctc 3135

Db   1184 AGAGCAGAACTCCTGACTACCATTCTATGACTGATCAAAAGACTAATAGTTAAAAACCTC 1243

Qy   3136 agcaggccttgttcacgatatgcag---aaaaagtgctgcagtttagatacctctgggaa 3192

Db   1244 AGCAGGCCTTGTTCACGATATGCAGAAAAAAAAAGTGCTGCAGTTTAGATACCTCT-GGAA 1302

Qy   3193 cttttccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaatatattt 3252

Db   1303 TTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTACATTTCTAAGAATATATTT 1362

Qy   3253 ctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagatttt 3312

Db   1363 CTTGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGGGCCCAAGATTTT 1422

Qy   3313 ttttt------------caataacagaccagcttctttttcctgcagttacaaatgtaa 3359

Db   1423 TTTTTTTTTTTTTTTTTCAAATAACAGACCAGCTTCTTTTTCTTGCAGTTACAGATGTAA 1482

Qy   3360 tttctttttttttttgttgtcaaacataaggtaccaaat---atgcaataaattgttttg 3416

Db   1483 TTTC-----CTTTTTGTTGTCAAACATAAGGTACCAAATATGATGCAATAAATTGTTTTG 1537


RESULT  13
AK023402
LOCUS       AK023402                1985 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ13340 fis, clone OVARC1001942, weakly similar
            to N-TERMINAL ACETYLTRANSFERASE 1 (EC 2.3.1.88).
ACCESSION   AK023402
VERSION     AK023402.1  GI:10435324
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_lib:OVARC1
            clone:OVARC1001942.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1985)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1985
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="OVARC1001942"
                     /tissue_type="ovary, tumor tissue"
                     /clone_lib="OVARC1"
                     /note="cloning vector: pME18SFL3"
     CDS             280. .1860
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB14562.1"
                     /db_xref="GI:10435325"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
```

```
                HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                HEIERKSLMTSLTFIHTV"
BASE COUNT      652 a    349 c    492 g    492 t
ORIGIN


  Query Match           31.6%;  Score 1079.6;  DB 9;  Length 1985;
  Best Local Similarity  91.5%;  Pred. No. 6.2e-206;
  Matches 1171;  Conservative    0;  Mismatches   94;  Indels   15;  Gaps    2;

Qy      27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81
           || ||||  || |||||||||||||||||||||||     ||||||||||||||||||| |
Db     710 GACCTGCGCAGAGAGCATCATGGATTGGTTATGCTATTGCTTACCATTTATTAGAAGATT 769

Qy      82 atgaaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgata 141
           ||||||||||||||| |||||||||| |||||||||| |||||||| ||||| |
Db     770 ATGAAATGGCAGCAAAGATTTTAGAAGAATTTAGGAAAACACAACAGACATCCCCTGACA 829

Qy     142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201
           | |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     830 AGGTGGATTATGAATATAGTGAACTACTCTTATATCAGAATCAAGTTCTTCGGGAAGCAG 889

Qy     202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261
           |||| ||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     890 GTCTCTATAGAGAAGCTTTGGAACATCTTTGTACCTATGAAAAGCAGATTTGTGATAAAC 949

Qy     262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321
           ||||||| |||||||||||||||||||||||||||||  | ||||||||||||||||||
Db     950 TTGCTGTAGAAGAAACCAAAGGGGAACTTCTGTTGCAACTATGTCGTTTGGAAGATGCTG 1009

Qy     322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381
           | || |||||||||||||| |||||||| |||||||||| ||||||||||||||||||||
Db    1010 CAGATGTTTATAGAGGATTGCAAGAGAGAAATCCTGAAAACTGGGCCTATTACAAAGGCT 1069

Qy     382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441
           | |||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1070 TGGAAAAAGCACTCAAGCCAGCTAATATGTTAGAACGGCTAAAAATTTATGAGGAAGCCT 1129

Qy     442 ggactaaataccccaggggactcgtgccaagaaagctgcccttaaacttttatctggag 501
           |||||||||| |||||||||| |||||||||| |||||| |||||||||||||||| |
Db    1130 GGACTAAATATCCCAGGGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTG 1189

Qy     502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561
           |||||||||| || |||||||||| |||||||||||||||||||||||||| || ||||
Db    1190 AGAAGTTTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGCCCAC 1249

Qy     562 ctgtcttcaataccttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621
           | |||||||||| || || || ||||||| ||| ||||| |||||||||||| ||||||
Db    1250 CAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGTGGCAATCATAGAAG 1309

Qy     622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681
           |  |||||||| |||||||||| |||||||||| || ||  ||||||||||||||||||
Db    1310 AGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATG 1369

Qy     682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741
           |||||||||||||||| ||||||||||||||||||||||||||||| |||||||| ||||
Db    1370 ATGGAAAGGAGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACATT 1429

Qy     742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801
           |||| |||||||||||||||| |||||| |||| |||||||||||||| ||||||||||
Db    1430 ATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTA 1489

Qy     802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861
           |||| ||||| |||||||||||||| || ||||||||||||||||||||||||| ||||
Db    1490 CACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATA 1549

Qy     862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921
           |||||||||||||| |||||||||||||| |||||||| ||||||||||||||||||||
Db    1550 TTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGATTTA 1609

Qy     922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981
           | || ||||| |||||||||||||||| ||||||||||||||||||||||| ||||||||||
Db    1610 TCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAA 1669

Qy     982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041
           |||| || |||||||| |||||||||||| |||||||||||||  ||||||||||||||
Db    1670 TGTGCTCAAAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAGT 1729

Qy    1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101
           | |||||||||||| ||||| ||||| ||||| || |||||||||| ||||||||||||
Db    1730 GCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAG 1789

Qy    1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
           |||||||||||||||||| |||||||||  ||||||| ||||||||||||||||||||
Db    1790 CACTTAAGAAATGTCATGAGATTGAGAGA----------AAATCACTGATGACCAGTTTG 1839

Qy    1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1840 ACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTTATTAAAAC 1899

Qy    1222 tagaagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgaga 1281
           |||||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||| ||||
Db    1900 TAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAGCAAGAATTGCTATAGAGA 1959

Qy    1282 tctatttgaagcttcatgac 1301
           ||||||||||||||||||||
Db    1960 TCTATTTGAAGCTTCATGAC 1979
```

RESULT 14
AC097376
LOCUS AC097376 145395 bp DNA linear PRI 03-JAN-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-83A24, complete sequence.
ACCESSION AC097376 AC018606
VERSION AC097376.3 GI:17149782
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145395)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145395)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108,.USA
REFERENCE 3 (bases 1 to 145395)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 145395)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 29, 2001 this sequence version replaced gi:16596615.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu

```
          ---------- Project Information ----------
          Center project name: H_NH0083A24
          Drafting center: WIBR
          ----------.

FEATURES             Location/Qualifiers
     source          1. .145395
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-83A24"
BASE COUNT    40225 a  27793 c  29529 g  47848 t
ORIGIN
```

Query Match 30.4%; Score 1038.4; DB 9; Length 145395;
Best Local Similarity 90.1%; Pred. No. 6.4e-198;
Matches 1315; Conservative 0; Mismatches 96; Indels 49; Gaps 17;

```
Qy   1983 cttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgtaaa 2042
          | | |||| || ||||| || |||||||||||||| |||||||||||| ||||||||||
Db  93961 CCTTAGACATGTATGGAGGTATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAA 94020

Qy   2043 gaagctgccgaagcctacagagcaagttgtcataagctttttcccttatgctttggctttc 2102
          |||||||| |||   || ||||||| ||||||||||||||||||||||||||||||||||||
Db  94021 GAAGCTGCTGAAATTTATAGAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTC 94080

Qy   2103 atgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcagaa 2162
          |||||||||||||| |||||||||||||||||||| || ||||||||||||||||||||
Db  94081 ATGCCTCCTGGATATGAAGAGGATATGAAGATCACAGTTAATGGAGATAGTTCTGCAGAA 94140

Qy   2163 acggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgactttgg 2222
          | ||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db  94141 GCTGAAGAACTGGCCAATGAAATTTGAACATCACTAAACAAGCAAATGGAATGACTTTGG 94200

Qy   2223 accatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacacag 2282
          |||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Db  94201 ACCATATCTAGTATATAATATTTTTGTCACGCACCTGCTGCATTGCTCTAACTTACACAG 94260

Qy   2283 aatgagaggagtaaatgttcttgccttcaaata--gtcttacgttttttatcctgctg-a 2339
          |||||||||||||||||||||||||||||||  || ||||||||||||||||||||| |
Db  94261 AATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTGTTTTACGTTTTTTATCCTGCTGAA 94320

Qy   2340 aaactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaa 2399
          ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||| ||
Db  94321 AAAGTATATATAAAATATCTAACATTACAGGATAGAGGTTCAGTTTCTTAAAAAATT-AA 94379

Qy   2400 agctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccc 2459
          |||||||||||||||| ||||      |||||| |  || |||||||||||| ||  |||
Db  94380 AGCTGCTAAAATTGAGTGGTT------AAAAAAGATACCTTATCCTATTCCTCCCCACCC 94433

Qy   2460 ttcccatgtttttaactaatttatatataaaatctggaggctataacagctaacatagcagg 2519
            ||    |||| |||||||||||||||||||||||||||| | ||||||||| ||||||
Db  94434 ACCCATGTTTTTAAACTAATTTATATAAAATCTGGAGGCTGTTACAGCTAACAAAGCAGG 94493

Qy   2520 tgtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagc 2579
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94494 TGTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTACTATATCTCAGACAGC 94553

Qy   2580 ataaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggct 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94554 ----ATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCT 94609

Qy   2640 ttagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtg 2699
          |||||||||| |||||| |||||||||||||| |   |||||||||||||||||||||||
Db  94610 TTAGCATCTGACTGATTTTGTTACGGGGTTGAT-TCTGACCATAGGAAGTATGCAATGTG 94668

Qy   2700 aatcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacat 2759
          ||||||| | |||||||    |||||||| |||||||| |||||||| ||||||||| ||
Db  94669 AATCACTATTTACAGAGAAACCTACAACAGATGCTTGATGTTGTAGAAACTGGGACATAT 94728

Qy   2760 agctaccaagc-ggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttcca 2818
          || ||||||||   ||||   |||||||| |||||||||| ||||    ||||||||||
Db  94729 AGATACCAAGCAAAATTATAAGAAACCTATAAGGTGTTCAATACG---CTTGTGTTTCCA 94785

Qy   2819 aaattcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaacc 2878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  94786 AAATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACTGAAGGAACC 94845

Qy   2879 agttggaacaatctcaatt-taactaaaacttgaagaactaaaataacaatgcaaaccttt 2937
          ||||| |||| |||||||| |||||||||||||||||||||||||| ||||||||||||||
Db  94846 AGTTGTAACAGTCTCAATTTTAACTAAAACTTGAAGAACTAAAACAACAATGCAAACCTT 94905

Qy   2938 tatcattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatg 2997
          |     |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94906 TCAGCATTGTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATG 94965

Qy   2998 tggcaccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactct 3057
          |||||||||||||||| ||||||||||||||||||  |||||||||||| ||| |||||||
Db  94966 TGGCACCAACTAATTA-GCAAGCATGAATTTTTCACCCAAGAGTGAAAAAAGGAAAATCT 95024

Qy   3058 accatggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagag 3117
          |||||||||||||||| | ||||||||||||||||||||||||||||||||||| || ||
Db  95025 ACCATGGCTTGAAGTTA-AAGAGCAGAACTCCTGACTACCATTCTATGACTGATCAAAAG 95083

Qy   3118 actaatatct-aaaacctcagcaggccttgttcacgatatgcag---aaaaagtgctgca 3173
          |||||||  | |||||||||||||||||||||||||||||||   ||||||||||||||
Db  95084 ACTAATAGTTAAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAAAAGTGCTGCA 95143

Qy   3174 gtttagatacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagccctt 3233
          |||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||
Db  95144 GTTTAGATACCTCT-GGAATTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTA 95202

Qy   3234 catttctaagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaattt 3293
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  95203 CATTTCTAAGAATATATTTCTTGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTT 95262

Qy   3294 ttaaagggcccaagatttttttt----------caataacagaccagcttctttt 3339
          |||||||||||||||||||||||          ||||||||||||||||||||||
Db  95263 TTAAAGGGCCCAAGATTTTTTTTTTTTTTTTTTTTTTTCAAATAACAGACCAGCTTCTTTT 95322

Qy   3340 tcctgcagttacaaatgtaatttctttttttttttgttgtcaaacataaggtaccaaat- 3398
          || |||||||||| ||||||||||      ||||||||||||||||||||||||||||||
Db  95323 TCTTGCAGTTACAGATGTAATTTC-----CTTTTTGTTGTCAAACATAAGGTACCAAATA 95377

Qy   3399 --atgcaataaattgttttg 3416
            ||||||||||||||||||
Db  95378 TGATGCAATAAATTGTTTTG 95397
```

```
RESULT 15
AC102860
LOCUS       AC102860               66729 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP24-511H12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC102860
VERSION     AC102860.1  GI:17060970
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 66729)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-511H12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66729)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
```

```
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19968
                Center clone name: 511_H_12
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      667: contig of 667 bp in length
            *      668  767: gap of      100 bp
            *      768     1462: contig of 695 bp in length
            *     1463 1562: gap of      100 bp
            *     1563     2258: contig of 696 bp in length
            *     2259 2358: gap of      100 bp
            *     2359     3065: contig of 707 bp in length
            *     3066 3165: gap of      100 bp
            *     3166     3842: contig of 677 bp in length
            *     3843 3942: gap of      100 bp
            *     3943     4599: contig of 657 bp in length
            *     4600 4699: gap of      100 bp
            *     4700     5378: contig of 679 bp in length
            *     5379 5478: gap of      100 bp
            *     5479     6178: contig of 700 bp in length
            *     6179 6278: gap of      100 bp
            *     6279     6988: contig of 710 bp in length
            *     6989 7088: gap of      100 bp
            *     7089     7796: contig of 708 bp in length
            *     7797 7896: gap of      100 bp
            *     7897     8572: contig of 676 bp in length
            *     8573 8672: gap of      100 bp
            *     8673     9382: contig of 710 bp in length
            *     9383 9482: gap of      100 bp
            *     9483    10183: contig of 701 bp in length
            *    10184 10283: gap of      100 bp
            *    10284    10984: contig of 701 bp in length
            *    10985 11084: gap of      100 bp
            *    11085    11788: contig of 704 bp in length
            *    11789 11888: gap of      100 bp
            *    11889    12582: contig of 694 bp in length
            *    12583 12682: gap of      100 bp
            *    12683    13381: contig of 699 bp in length
            *    13382 13481: gap of      100 bp
            *    13482    14182: contig of 701 bp in length
            *    14183 14282: gap of      100 bp
            *    14283    14955: contig of 673 bp in length
            *    14956 15055: gap of      100 bp
            *    15056    15754: contig of 699 bp in length
            *    15755 15854: gap of      100 bp
            *    15855    16558: contig of 704 bp in length
            *    16559 16658: gap of      100 bp
            *    16659    17346: contig of 688 bp in length
            *    17347 17446: gap of      100 bp
            *    17447    18139: contig of 693 bp in length
            *    18140 18239: gap of      100 bp
            *    18240    18954: contig of 715 bp in length
            *    18955 19054: gap of      100 bp
            *    19055    19743: contig of 689 bp in length
            *    19744 19843: gap of      100 bp
            *    19844    20534: contig of 691 bp in length
            *    20535 20634: gap of      100 bp
            *    20635    21329: contig of 695 bp in length
            *    21330 21429: gap of      100 bp
            *    21430    22118: contig of 689 bp in length
            *    22119 22218: gap of      100 bp
            *    22219    22925: contig of 707 bp in length
            *    22926 23025: gap of      100 bp
            *    23026    23712: contig of 687 bp in length
            *    23713 23812: gap of      100 bp
            *    23813    24526: contig of 714 bp in length
            *    24527 24626: gap of      100 bp
            *    24627    25332: contig of 706 bp in length
            *    25333 25432: gap of      100 bp
            *    25433    26124: contig of 692 bp in length
            *    26125 26224: gap of      100 bp
            *    26225    26928: contig of 704 bp in length
            *    26929 27028: gap of      100 bp
            *    27029    27743: contig of 715 bp in length
            *    27744 27843: gap of      100 bp
            *    27844    28532: contig of 689 bp in length
            *    28533 28632: gap of      100 bp
            *    28633    29328: contig of 696 bp in length
            *    29329 29428: gap of      100 bp
            *    29429    30116: contig of 688 bp in length
            *    30117 30216: gap of      100 bp
            *    30217    30926: contig of 710 bp in length
            *    30927 31026: gap of      100 bp
            *    31027    31728: contig of 702 bp in length
            *    31729 31828: gap of      100 bp
            *    31829    32519: contig of 691 bp in length
            *    32520 32619: gap of      100 bp
            *    32620    33313: contig of 694 bp in length
            *    33314 33413: gap of      100 bp
            *    33414    34116: contig of 703 bp in length
            *    34117 34216: gap of      100 bp
            *    34217    34884: contig of 668 bp in length
            *    34885 34984: gap of      100 bp
            *    34985    35686: contig of 702 bp in length
            *    35687 35786: gap of      100 bp
            *    35787    36435: contig of 649 bp in length
            *    36436 36535: gap of      100 bp
            *    36536    37233: contig of 698 bp in length
            *    37234 37333: gap of      100 bp
            *    37334    38013: contig of 680 bp in length
            *    38014 38113: gap of      100 bp
            *    38114    38820: contig of 707 bp in length
            *    38821 38920: gap of      100 bp
            *    38921    39614: contig of 694 bp in length
            *    39615 39714: gap of      100 bp
            *    39715    40417: contig of 703 bp in length
            *    40418 40517: gap of      100 bp
            *    40518    41219: contig of 702 bp in length
            *    41220 41319: gap of      100 bp
            *    41320    42018: contig of 699 bp in length
            *    42019 42118: gap of      100 bp
            *    42119    42838: contig of 720 bp in length
            *    42839 42938: gap of      100 bp
            *    42939    43639: contig of 701 bp in length
```

```
*    43640 43739: gap of      100 bp
*    43740    44442: contig of 703 bp in length
*    44443 44542: gap of      100 bp
*    44543    45226: contig of 684 bp in length
*    45227 45326: gap of      100 bp
*    45327    46009: contig of 683 bp in length
*    46010 46109: gap of      100 bp
*    46110    46803: contig of 694 bp in length
*    46804 46903: gap of      100 bp
*    46904    47586: contig of 683 bp in length
*    47587 47686: gap of      100 bp
*    47687    48389: contig of 703 bp in length
*    48390 48489: gap of      100 bp
*    48490    49189: contig of 700 bp in length
*    49190 49289: gap of      100 bp
*    49290    50004: contig of 715 bp in length
*    50005 50104: gap of      100 bp
*    50105    50800: contig of 696 bp in length
*    50801 50900: gap of      100 bp
*    50901    51592: contig of 692 bp in length
*    51593 51692: gap of      100 bp
*    51693    52331: contig of 639 bp in length
*    52332 52431: gap of      100 bp
*    52432    53123: contig of 692 bp in length
*    53124 53223: gap of      100 bp
*    53224    53948: contig of 725 bp in length
*    53949 54048: gap of      100 bp
*    54049    54758: contig of 710 bp in length
*    54759 54858: gap of      100 bp
```

Query Match 18.3%; Score 626.6; DB 2; Length 66729;
Best Local Similarity 98.7%; Pred. No. 1.6e-115;
Matches 684; Conservative 0; Mismatches 4; Indels 5; Gaps 5;

```
Qy   2655 ttctgttacggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacag 2714
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23837 TCCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATGTGAATCACTGTGTACAG 23896

Qy   2715 agccgtctacaacacatgcttgacgttgtagagactgggacacatagctaccaagcggat 2774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23897 AGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACACATAGCTACCAAGCGGAT 23956

Qy   2775 taagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatg 2834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23957 TAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAAATTCACTGTACATG 24016

Qy   2835 atcagtttggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctca 2894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24017 ATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAGTTGGAACAATCTCA 24076

Qy   2895 atttaactaaaacttgaagaactaaaataacaatgcaaacctttatcattgttttggcca 2954
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  24077 ATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTTTAGCATTGTTTTGGCCA 24136

Qy   2955 aacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaattat 3014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24137 AACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGGCACCAACTAATTA- 24195

Qy   3015 gcaagcatgaatttttcacctgagagtgaaaaaagaaaactctaccatggcttgaagtta 3074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24196 GCAAGCATGAATTTTTCACCTGAGAGTGAAAAAAGAAAACTCTACCATGGCTTGAAGTTA 24255

Qy   3075 caggagcagaactcctgactaccattctatgactgatgaagagactaatatctaaaacct 3134
          || |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  24256 CA-GAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAGACTAATAGCTAAAACCT 24314

Qy   3135 cagcaggccttgttcacgatatgcagaaaaagtgctgcagtttagatacctctgggaact 3194
          |||||||||||||||||||||||| |||||||||||||||||||||||||| ||||||
Db  24315 CAGCAGGCCTTGTTCACGATATGCAG-AAAAGTGCTGCAGTTTAGATACCTCT-GGAACT 24372

Qy   3195 tttccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaata-tatttc 3253
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  24373 TTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTCTAAGAATANTATTTC 24432

Qy   3254 tcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagattttt 3313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24433 TCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGGGCCCAAGATTTTT 24492

Qy   3314 ttttcaataacagaccagcttctttttcctgca 3346
          |||||||||||||||||||||||||| |||||
Db  24493 TTTTCAATAACAGACCAGCTTCTTTTTTCTGCA 24525
```

Search completed: July 20, 2002, 02:41:30
Job time: 14548 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 20, 2002, 00:31:22 ; Search time 356.49 Seconds
(without alignments)
16461.650 Million cell updates/sec

Title: US-09-836-410A-2
Perfect score: 3418
Sequence: 1 caagtaacacccgcaagatg.....atgcaataaattgttttggg 3418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3418	100.0	3418	24	AAD22687	Tubedown-1 (tbdn-1
	2	3418	100.0	3418	24	AAH77156	Human tubedown-1 c
c	3	3418	100.0	3418	24	AAH77158	Human tubedown-1 b
	4	1502.2	43.9	2477	23	AAS71925	DNA encoding novel
c	5	1413	41.3	1413	24	AAH77157	Human tubedown-1 b
	6	1210.2	35.4	1802	22	AAH16408	Human cDNA sequenc
	7	1126	32.9	2859	22	AAH14477	Human cDNA sequenc
	8	1079.6	31.6	1985	22	AAH16424	Human cDNA sequenc
c	9	617.8	18.1	710	22	AAH81664	Human differential
	10	616.2	18.0	790	22	AAH06489	Human cDNA clone (
	11	565	16.5	1225	20	AAX99092	Human cancer cell
	12	555.6	16.3	781	20	AAZ15705	Human gene express
	13	555.6	16.3	781	20	AAX99053	Human validated ca
	14	475.6	13.9	764	20	AAZ15983	Human gene express
	15	475.6	13.9	764	20	AAX98777	Human validated ca
	16	398.4	11.7	774	20	AAX98889	Human validated ca
c	17	343.6	10.1	488	22	AAH12222	Human cDNA clone (
	18	342.6	10.0	408	22	AAS37727	Novel human diagno
	19	340	9.9	408	22	AAS37726	Novel human diagno
	20	339.8	9.9	402	22	AAS37350	Novel human diagno
	21	310.6	9.1	773	20	AAX99063	Human validated ca
	22	281	8.2	404	22	AAF66446	Novel human polynu
	23	276.6	8.1	727	22	AAI96842	Human neuroblastom
	24	268	7.8	300	20	AAX98340	Human cancer cell
	25	258.4	7.6	300	20	AAZ12696	Human gene express
	26	258.4	7.6	300	20	AAX98342	Human cancer cell
	27	242.6	7.1	399	20	AAV89559	EST clone CR392.
c	28	239	7.0	255	22	AAH81986	Rat differential t
	29	226	6.6	288	22	AAH81733	Human differential
	30	215.4	6.3	255	22	AAH82031	Rat differential t
	31	214.4	6.3	300	20	AAX98346	Human cancer cell
	32	212.8	6.2	300	20	AAZ12698	Human gene express
	33	188	5.5	653	22	AAH07206	Human cDNA clone (
	34	163.8	4.8	802	21	AAA02414	Human colon cancer
	35	149.4	4.4	297	21	AAC21231	Human secreted pro
	36	73.4	2.1	2703	10	AAN90541	DNA encoding N-alp
	37	73.4	2.1	2724	12	AAQ12226	N-alpha-acetyltran
	38	57.2	1.7	654	23	AAS69553	DNA encoding novel
	39	57.2	1.7	654	23	AAS71153	DNA encoding novel
	40	57.2	1.7	654	23	AAS75467	DNA encoding novel
	41	57	1.7	450	22	AAL00120	Human reproductive
c	42	56	1.6	887	22	AAI94064	Human neuroblastom
c	43	55.6	1.6	14006	24	ABL33958	Human immune syste
c	44	54.2	1.6	6171	24	ABL32788	Human immune syste
c	45	53.8	1.6	12007	24	ABL32717	Human immune syste

ALIGNMENTS

RESULT 1
AAD22687
ID AAD22687 standard; cDNA; 3418 BP.
XX
AC AAD22687;
XX
DT 26-FEB-2002 (first entry)
XX
DE Tubedown-1 (tbdn-1) protein encoding cDNA.
XX
KW Tubedown-1 protein; tbdn-1; ophthalmological; cytostatic; vulnerary;
KW cerebroprotective; angiogenesis inhibitor; ocular neovascularisation;
KW retinal disease; diabetic retinopathy; retinopathy of prematurity;
KW primary hyperplastic vitreous; macular degeneration; trauma; stroke;
KW haemorrhagic shock; arthritis; arteriosclerosis; delayed wound healing;
KW angiofibroma; granulation; nonunion fracture; retrolental fibroplasia;
KW solid tumour growth; chronic glaucoma; sickle cell retinopathy; cancer;
KW burn; scar; corneal neovascularisation; rubeosis iritis; uveitis;
KW gene therapy; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 18..2189
FT /*tag= a
FT /product= "Tubedown-1 protein #2"
FT CDS 57..2189
FT /*tag= b
FT /product= "Tubedown-1 protein #3"
FT CDS 87..2189
FT /*tag= c
FT /product= "Tubedown-1 protein #4"

```
FT   CDS             408..2189
FT                   /*tag=  d
FT                   /product= "Tubedown-1 protein #1"
FT   misc_feature    408..2186
FT                   /note= "This region is specifically claimed as
FT                   SEQ ID NO: 6 in claim 5 of the specification"
XX
PN   WO200179506-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12548.
XX
PR   17-APR-2000; 2000US-197977P.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-026032/03.
DR   P-PSDB; AAE13589, AAE13590, AAE13591, AAE13592.
XX
PT   Novel tubedown-1 protein comprising anti-angiogenic activity is useful
PT   for treating angiogenesis-associated disease related to ocular
PT   neovascularization, e.g., diabetic retinopathy, retinopathy of
PT   prematurity -
XX
PS   Claim 10; Page 56-58; 85pp; English.
XX
CC   The present invention relates to tubedown-1 (tbdn-1) proteins and
CC   their corresponding cDNAs. Tbdn-1 proteins having anti-angiogenic
CC   activity are associated with acetyl transferase activity. They
CC   regulate endothelial differentiation through protein acetylation,
CC   DNA-binding or by interacting with and/or acetylating other protein
CC   targets important for endothelial differentiation. In normal adult
CC   eyes, tbdn-1 is expressed highly in the corneal endothelium proper
CC   and in the vascular endothelium of the limbus and retina. Tbdn-1
CC   proteins are useful for preventing, treating, inhibiting or delaying
CC   the onset of angiogenesis-associated disease related to ocular
CC   neovascularisation, e.g., a retinal disease, such as preferably
CC   diabetic retinopathy or retinopathy of prematurity, or primary
CC   hyperplastic vitreous, macular degeneration and any other conditions
CC   involving ocular neovascularisation. Tbdn-1 proteins are also useful
CC   for treating any pathological neovascularisation condition such as
CC   head trauma, spinal trauma, stroke, haemorrhagic shock, arthritis,
CC   arteriosclerosis, angiofibroma, delayed wound healing, granulations,
CC   cancer, burns, scars, nonunion fractures, retrolental fibroplasia,
CC   solid tumour growth. Proteins of the invention are also useful for
CC   treating ocular neovascularisation conditions such as chronic glaucoma,
CC   sickle cell retinopathy, corneal neovascularisation, rubeosis iritis,
CC   uveitis, neovascularisation of the optic nerve. Sequences of the
CC   invention are also used in gene therapy. The present sequence is
CC   a cDNA encoding tubedown-1 proteins.
XX
SQ   Sequence 3418 BP; 1157 A; 604 C; 704 G; 953 T; 0 other;

  Query Match             100.0%;  Score 3418;  DB 24;  Length 3418;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
Db     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaagggaacttctgttgcagt 300
Db     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaagggaacttctgttgcagt 300

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
Db     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
Db     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
Db     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480

Qy     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
Db     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540

Qy     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
Db     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
Db     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
Db     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
Db     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
Db     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
Db     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
Db     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
Db     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020

Qy    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080

Qy    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140

Qy    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200

Qy    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260
Db    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260

Qy    1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
```

```
Db  1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Qy  1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Db  1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Qy  1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Db  1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Qy  1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
Db  1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
Qy  1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Db  1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Qy  1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Db  1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Qy  1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Db  1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Qy  1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Db  1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Qy  1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Db  1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Qy  1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttttctgaaaagga 1860
Db  1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttttctgaaaagga 1860
Qy  1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Db  1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Qy  1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Db  1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Qy  1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Db  1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Qy  2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100
Db  2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100
Qy  2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
Db  2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
Qy  2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Db  2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Qy  2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Db  2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Qy  2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgctgaa 2340
Db  2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgttttttatcctgctgaa 2340
Qy  2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Db  2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Qy  2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Db  2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Qy  2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Db  2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Qy  2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
Db  2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
Qy  2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Db  2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Qy  2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
Db  2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
Qy  2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
Db  2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
Qy  2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
Db  2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
Qy  2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
Db  2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
Qy  2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
Db  2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
Qy  2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
Db  2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
Qy  3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
Db  3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
Qy  3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
Db  3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
Qy  3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
Db  3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
Qy  3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
Db  3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
Qy  3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
Db  3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
Qy  3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360
Db  3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360
Qy  3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
Db  3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
```

RESULT 2
AAH77156

```
ID   AAH77156 standard; cDNA; 3418 BP.
XX
AC   AAH77156;
XX
DT   21-JAN-2002  (first entry)
XX
DE   Human tubedown-1 cDNA.
XX
KW   Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW   bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             408..2189
FT                   /*tag=  a
FT                   /product= "tubedown-1"
XX
PN   WO200179505-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12435.
XX
PR   17-APR-2000; 2000US-197977P.
PR   17-APR-2001; 2001US-0836410.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-017618/02.
DR   P-PSDB; AAG77907.
XX
PT   Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT   overexpression of tubedown-1 protein and are useful to treat
PT   osteosarcoma and Ewing's Sarcoma family of tumours  -
XX
PS   Claim 1; Page 36-38; 56pp; English.
XX
CC   The sequence represents a new human gene, tubedown-1 (tbdn-1). The
CC   invention relates to a novel isolated nucleic acid of the tubedown-1
CC   gene, and antisense nucleic acids to tbdn-1. The polynucleotides and
CC   protein of the invention have cytostatic and osteopathic activity. The
CC   polynucleotides of the invention may be used in antisense-therapy/gene
CC   therapy. They are useful in the treatment of bone tumours, especially
CC   osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC   invention may also be useful for the prevention of metastases from these
CC   types of tumours, either alone or in combination with radiotherapy and/or
CC   chemotherapeutic agents.
XX
SQ   Sequence 3418 BP; 1157 A; 604 C; 704 G; 953 T; 0 other;
```

Query Match 100.0%; Score 3418; DB 24; Length 3418;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
Db     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
Db     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
Db     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
Db     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420

Qy     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
Db     421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480

Qy     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
Db     481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540

Qy     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
Db     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
Db     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
Db     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780
Db     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaataca 780

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
Db     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
Db     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
Db     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
Db     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020

Qy    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080

Qy    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140

Qy    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200

Qy    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260
Db    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260

Qy    1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Db    1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
```

```
Qy   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380

Qy   1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440

Qy   1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500

Qy   1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560

Qy   1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620

Qy   1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680

Qy   1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740

Qy   1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800

Qy   1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttctctgaaaagga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttctctgaaaagga 1860

Qy   1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920

Qy   1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980

Qy   1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040

Qy   2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 aagaagctgccgaagcctacagagcaagttgtcataagcttttcccttatgctttggctt 2100

Qy   2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160

Qy   2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220

Qy   2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280

Qy   2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgtttttatcctgctgaa 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgtttttatcctgctgaa 2340

Qy   2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400

Qy   2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460

Qy   2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520

Qy   2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580

Qy   2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640

Qy   2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700

Qy   2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760

Qy   2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820

Qy   2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880

Qy   2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940

Qy   2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000

Qy   3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060

Qy   3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120

Qy   3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180

Qy   3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240

Qy   3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300

Qy   3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360

Qy   3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
```

RESULT 3
AAH77158/c
ID AAH77158 standard; cDNA; 3418 BP.
XX
AC AAH77158;

XX
DT 21-JAN-2002 (first entry)
XX
DE Human tubedown-1 base pairs 3418-1 antisense cDNA.
XX
KW Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200179505-A2.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-US12435.
XX
PR 17-APR-2000; 2000US-197977P.
PR 17-APR-2001; 2001US-0836410.
XX
PA (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI Gendron RL, Paradis H;
XX
DR WPI; 2002-017618/02.
XX
PT Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT overexpression of tubedown-1 protein and are useful to treat
PT osteosarcoma and Ewing's Sarcoma family of tumours -
XX
PS Claim 7; Page 39-41; 56pp; English.
XX
CC The sequence represents tubedown-1 (tbdn-1) bases 3418-1 antisense cDNA.
CC The invention relates to a novel isolated nucleic acid of the tubedown-1
CC gene, and antisense nucleic acids to tbdn-1. The polynucleotides and
CC protein of the invention have cytostatic and osteopathic activity. The
CC polynucleotides of the invention may be used in antisense-therapy/gene
CC therapy. They are useful in the treatment of bone tumours, especially
CC osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC invention may also be useful for the prevention of metastases from these
CC types of tumours, either alone or in combination with radiotherapy and/or
CC chemotherapeutic agents.
XX
SQ Sequence 3418 BP; 953 A; 704 C; 604 G; 1157 T; 0 other;

Query Match 100.0%; Score 3418; DB 24; Length 3418;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db    3418 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 3359

Qy      61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db    3358 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 3299

Qy     121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db    3298 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 3239

Qy     181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
Db    3238 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 3179

Qy     241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
Db    3178 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 3119

Qy     301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
Db    3118 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 3059

Qy     361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
Db    3058 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 2999

Qy     421 taaaaatatatgaggaagcctggactaaatacccccaggggactcgtgccaagaaagctgc 480
Db    2998 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 2939

Qy     481 ccttaaacttttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
Db    2938 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 2879

Qy     541 atttcagcaagggctgtccacctgtcttcaataccttgaggtctttatacagagataaag 600
Db    2878 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 2819

Qy     601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
Db    2818 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 2759

Qy     661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
Db    2758 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 2699

Qy     721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaatataca 780
Db    2698 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 2639

Qy     781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
Db    2638 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 2579

Qy     841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
Db    2578 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 2519

Qy     901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaaccc 960
Db    2518 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 2459

Qy     961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
Db    2458 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 2399

Qy    1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
Db    2398 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 2339

Qy    1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
Db    2338 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 2279

Qy    1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
Db    2278 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 2219

Qy    1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccatttttacttcaaag 1260
Db    2218 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 2159

Qy    1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
Db    2158 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 2099

Qy    1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
Db    2098 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 2039

Qy    1381 ataaacaaagaagagctcaaaagaaagcccagattgaagaagagaaaaaaaatgccgaaa 1440
Db    2038 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATTGAAGAAGAGAAAAAAAATGCCGAAA 1979

Qy    1441 aagaaaagccgcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggag 1500
```

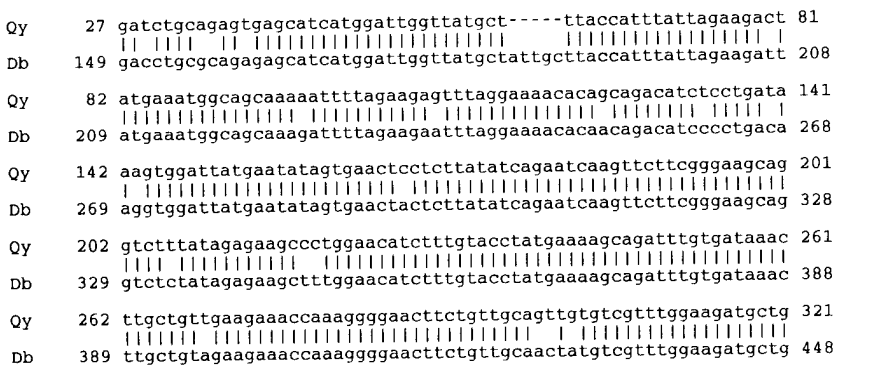
```
Db   1978 AAGAAAAGCCGCAACGGAATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAG 1919

Qy   1501 gccccaaagaagagcttatccctgagaaactggccaaggttgaaactccattggaagaag 1560
Db   1918 GCCCCAAAGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAGAAG 1859

Qy   1561 ctattaagtttttaacaccattgaagaacttggtgaagaacaagatagaaactcatcttt 1620
Db   1858 CTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATAGAAACTCATCTTT 1799

Qy   1621 ttgcctttgagatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagc 1680
Db   1798 TTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCTTTTGATGCTACAATCAGTAAAGC 1739

Qy   1681 gggcatttgctattgattctagtcatccctggcttcatgagtgcatgattcgactctttc 1740
Db   1738 GGGCATTTGCTATTGATTCTAGTCATCCCTGGCTTCATGAGTGCATGATTCGACTCTTTC 1679

Qy   1741 attctgtgtgtgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaa 1800
Db   1678 ATTCTGTGTGTGAAAGTAAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAA 1619

Qy   1801 tgaatcgtctttttggagcaacaaatccaaagaattttaatgaaaccttttctgaaaagga 1860
Db   1618 TGAATCGTCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAAGGA 1559

Qy   1861 attctgattcattgccacatagattatcagctgccaaaatggtatattatttagattctt 1920
Db   1558 ATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATATTATTTAGATTCTT 1499

Qy   1921 ctagtcaaaaacgagcaatagagctggcgacaacacttgatggatccctcaccaacagaa 1980
Db   1498 CTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACTTGATGGATCCCTCACCAACAGAA 1439

Qy   1981 accttcagacttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgta 2040
Db   1438 ACCTTCAGACTTGCATGGAAGTGTTGGAAGCCTTGTGTGATGGTAGCCTACGAGACTGTA 1379

Qy   2041 aagaagctgccgaagcctacagagcaagttgtcataagctttttcccttatgctttggctt 2100
Db   1378 AAGAAGCTGCCGAAGCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTT 1319

Qy   2101 tcatgcctcctggatacgaagaggatatgaagatcacagtgaacggagatagttctgcag 2160
Db   1318 TCATGCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTGCAG 1259

Qy   2161 aaacggaagaactggccaatgaaatctgaacatcattaaacaagcaaatggaatgacttt 2220
Db   1258 AAACGGAAGAACTGGCCAATGAAATCTGAACATCATTAAACAAGCAAATGGAATGACTTT 1199

Qy   2221 ggaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacac 2280
Db   1198 GGACCATATCTAGTGTATAATATTTTTGTCACGCACCTGCTGCATTGCTCTTACTTACAC 1139

Qy   2281 agaatgagaggagtaaatgttcttgccttcaaatagtcttacgtttttatcctgctgaa 2340
Db   1138 AGAATGAGAGGAGTAAATGTTCTTGCCTTCAAATAGTCTTACGTTTTTTATCCTGCTGAA 1079

Qy   2341 aactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaaaa 2400
Db   1078 AACTATATATAAAATATCTAACATTACAGGATATAGGTTCAGTTTCTTAAAAAATTAAAA 1019

Qy   2401 gctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttccct 2460
Db   1018 GCTGCTAAAATTGAGGGGTTTAAAAGAAAAAAAAATCCGTATCCTATTCCTACCTTCCCT 959

Qy   2461 tcccatgtttttaactaatttatataaaatctggaggctataacagctaacatagcaggt 2520
Db    958 TCCCATGTTTTTAACTAATTTATATAAAATCTGGAGGCTATAACAGCTAACATAGCAGGT 899

Qy   2521 gtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagacagca 2580
```

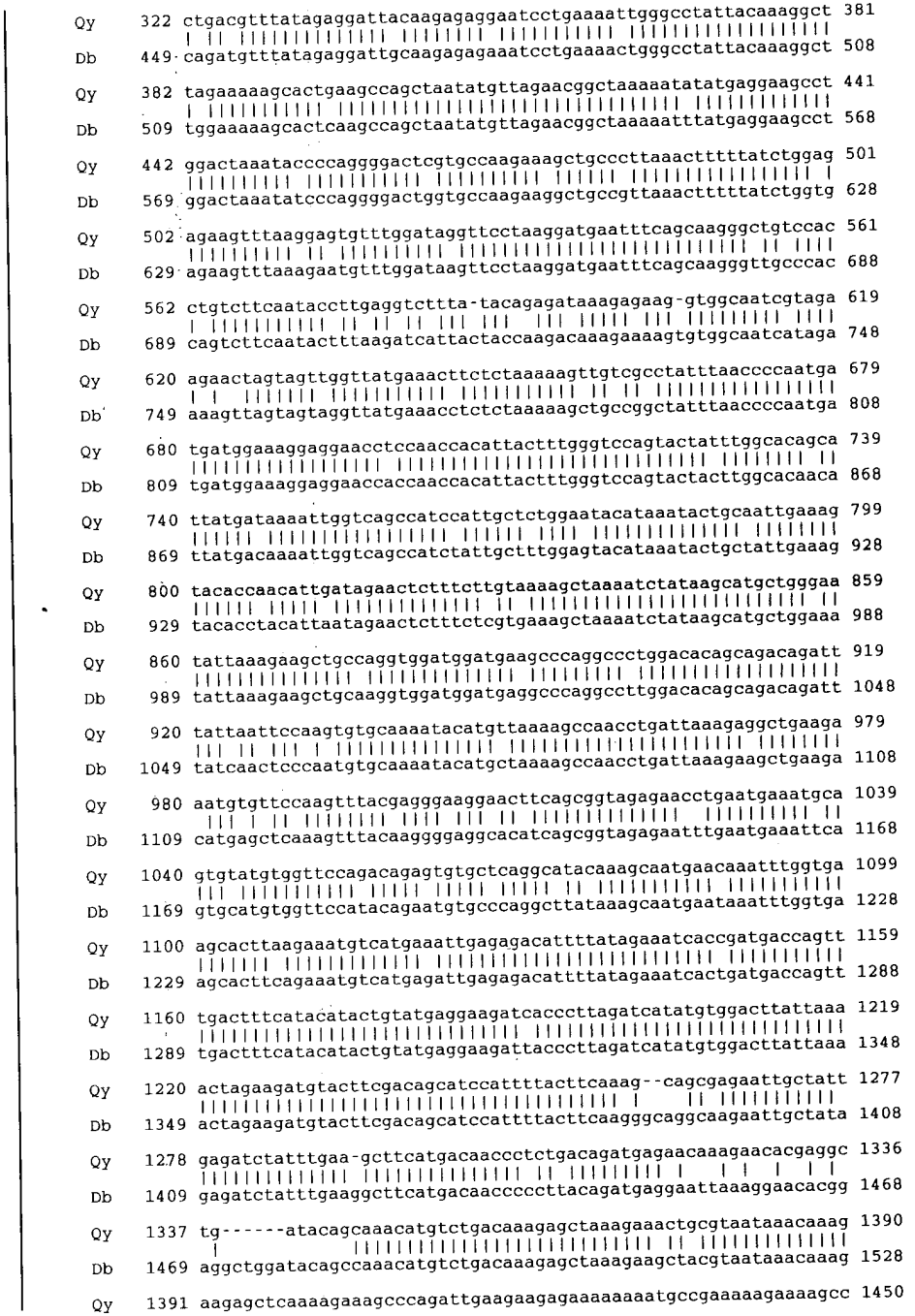
```
Db    898 GTGTGGCAGAAATATTACTTTAAATTTGTCTTGTGAGATTTTGCTATATCTCAGACAGCA 839

Qy   2581 taaataaatgctgttttagcactggattctttcactgagcacaaagagttgttggggctt 2640
Db    838 TAAATAAATGCTGTTTTAGCACTGGATTCTTTCACTGAGCACAAAGAGTTGTTGGGGCTT 779

Qy   2641 tagcatctgcctgattctgttacggggttggtgattgaccataggaagtatgcaatgtga 2700
Db    778 TAGCATCTGCCTGATTCTGTTACGGGGTTGGTGATTGACCATAGGAAGTATGCAATGTGA 719

Qy   2701 atcactgtgtacagagccgtctacaacacatgcttgacgttgtagagactgggacacata 2760
Db    718 ATCACTGTGTACAGAGCCGTCTACAACACATGCTTGACGTTGTAGAGACTGGGACACATA 659

Qy   2761 gctaccaagcggattaagtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaa 2820
Db    658 GCTACCAAGCGGATTAAGTGAAACCTAGAAGGTGTTCAGTACGTGTGTTGTGTTTCCAAA 599

Qy   2821 attcactgtacatgatcagtttggtgttcttgtaccacagtttttaaccgaaggaaccag 2880
Db    598 ATTCACTGTACATGATCAGTTTGGTGTTCTTGTACCACAGTTTTTAACCGAAGGAACCAG 539

Qy   2881 ttggaacaatctcaatttaactaaaacttgaagaactaaaataacaatgcaaacctttat 2940
Db    538 TTGGAACAATCTCAATTTAACTAAAACTTGAAGAACTAAAATAACAATGCAAACCTTTAT 479

Qy   2941 cattgttttggccaaacttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtgg 3000
Db    478 CATTGTTTTGGCCAAACTTGTTAAAACTGTAATGCAAGAACCAAATGCACTGTGATGTGG 419

Qy   3001 caccaactaattatgcaagcatgaatttttcacctgagagtgaaaaaagaaaactctacc 3060
Db    418 CACCAACTAATTATGCAAGCATGAATTTTTCACCTGAGAGTGAAAAAAGAAAACTCTACC 359

Qy   3061 atggcttgaagttacaggagcagaactcctgactaccattctatgactgatgaagagact 3120
Db    358 ATGGCTTGAAGTTACAGGAGCAGAACTCCTGACTACCATTCTATGACTGATGAAGAGACT 299

Qy   3121 aatatctaaaacctcagcaggccttgttcacgatatgcagaaaaagtgctgcagtttaga 3180
Db    298 AATATCTAAAACCTCAGCAGGCCTTGTTCACGATATGCAGAAAAAGTGCTGCAGTTTAGA 239

Qy   3181 tacctctgggaacttttccacagtgtcacaggtttgtaatacttgaagcccttcatttct 3240
Db    238 TACCTCTGGGAACTTTTCCACAGTGTCACAGGTTTGTAATACTTGAAGCCCTTCATTTCT 179

Qy   3241 aagaatatatttctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagg 3300
Db    178 AAGAATATATTTCTCGCTCAGTTGTTTCAGGCAAGCCCAAGACTTTGTAATTTTTAAAGG 119

Qy   3301 gcccaagatttttttttcaataacagaccagcttctttttcctgcagttacaaatgtaat 3360
Db    118 GCCCAAGATTTTTTTTTCAATAACAGACCAGCTTCTTTTTCCTGCAGTTACAAATGTAAT 59

Qy   3361 ttctttttttttttgttgtcaaacataaggtaccaaatatgcaataaattgttttggg 3418
Db     58 TTCTTTTTTTTTTTGTTGTCAAACATAAGGTACCAAATATGCAATAAATTGTTTTGGG 1
```

```
RESULT  4
AAS71925
ID   AAS71925 standard; cDNA; 2477 BP.
XX
AC   AAS71925;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7729.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
```

XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG07738.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 7729; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2477 BP; 859 A; 437 C; 522 G; 659 T; 0 other;

Query Match 43.9%; Score 1502.2; DB 23; Length 2477;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 2082; Conservative 0; Mismatches 218; Indels 279; Gaps 17;





```
Db   1529 aagagctcaaaagaaagcccagatagaagaagagaaaaaaaatgcagaaaaagaaaagca 1588

Qy   1451 gcaacggaatccgaaaaagaaaaaggatgatgatgacgaagaaattggaggccccaaaga 1510
Db   1589 gcagagaaatcagaaaaagaagaaggatgatgatgatgaggagataggaggtccaaaaga 1648

Qy   1511 agagcttatccctgagaaactggccaaggttgaaactccattggaagaagctattaagtt 1570
Db   1649 agaacttattccagagaaactggccaaggttgaaactccattggaagaagctattaaatt 1708

Qy   1571 tttaacaccattgaagaacttggtgaagaacaagatagaaactcatctttttgcctttga 1630
Db   1709 tttaacaccgttgaagaacttggtgaagaacaagatagagactcatctttttgcctttga 1768

Qy   1631 gatctactttaggaaagaaaagtttcttttgatgctacaatcagtaaagcgggcatttgc 1690
Db   1769 gatttactttaggaa---------------------------------------------- 1783

Qy   1691 tattgattctagtcatccctggcttcatgagtgcatgattcgactctttcattctgtgtg 1750

Db   1784 ---------------------------------------------------------- 1783

Qy   1751 tgaaagtaaagacttacccgaaacagttagaacagtattaaaacaagaaatgaatcgtct 1810

Db   1784 ---------------------------------------------------------- 1783

Qy   1811 ttttggagcaacaaatccaaagaattttaatgaaaccttttctgaaaaggaattctgattc 1870

Db   1784 ---------------------------------------------------------- 1783

Qy   1871 attgccacatagattatcagctgccaaaatggtatattatttagattcttctagtcaaaa 1930
Db   1784 ------------------agctgccaaaatggtatattacttagatccttctagtcagaa 1825

Qy   1931 acgagcaatagagctggcgacaacacttgatggatccctcaccaacagaaaccttcagac 1990
Db   1826 gcgagctatagagttggcaacaacacttgatgaatctctcactaacagaaacctccagac 1885

Qy   1991 ttgcatggaagtgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgc 2050
Db   1886 atgtatggaggtattggaagccttgtatgatggtagcctaggagactgtaaagaagctgc 1945

Qy   2051 cgaagcctacagagcaagttgtcataagctttt--cccttatgctttgg----ctttcat 2104
Db   1946 tgaaatttatagagcaaattgtccataagctttaccccttatgcctttggccttccatg 2005

Qy   2105 gcctcctggatacgaagaggatatgaagat-cacagtgaacggagatagtt-ctgcagaa 2162
Db   2006 ccccccggatatgaagaggatatgaagatccacagttaatggagatagttcctgcagaa 2065

Qy   2163 acggaagaactgg-ccaatgaaatctgaacatcattaaacaagcaaatggaatgactttg 2221
Db   2066 gctgaagaactggcccaatgaaatttgaacatcactaaacaagcaaatggaatgactttg 2125

Qy   2222 gaccatatctagtgtataatatttttgtcacgcacctgctgcattgctcttacttacaca 2281
Db   2126 gaccatatctagtatataatatttttgtcacgcacctgctgcattgctctaacttacaca 2185

Qy   2282 gaatgagaggagtaaatgttcttgccttcaaatagtc--ttacgttttttatcctgctg- 2338
Db   2186 gaatgagaggagtaaatgttcttgccttcaaatagtgttttacgttttttatcctgctga 2245

Qy   2339 aaaactatatataaaatatctaacattacaggatataggttcagtttcttaaaaaattaa 2398
Db   2246 aaaagtatatataaaatatctaacattacaggatagaggttcagtttcttaaaaaatt-a 2304

Qy   2399 aagctgctaaaattgaggggtttaaaagaaaaaaaaatccgtatcctattcctaccttcc 2458
Db   2305 aagctgctaaaattgagtggtt------aaaaaagataccttatcctattcctccccacc 2358

Qy   2459 cttcccatgtttttaa-ctaatttatataaaatctggaggctataacagctaacatagca 2517
Db   2359 cacccatgtttttaaacctaatttatataaaatctggaggctgttacagctaacaaagca 2418

Qy   2518 ggtgtgtggcagaaatattactttaaatttgtcttgtgagattttgctatatctcagac 2576
Db   2419 ggtgtgtggcagaaatattactttaaatttgtcttgtgagattttactatatctcagac 2477
```

RESULT 5
AAH77157/c
ID AAH77157 standard; cDNA; 1413 BP.
XX
AC AAH77157;
XX
DT 21-JAN-2002 (first entry)
XX
DE Human tubedown-1 base pairs 1413-1 antisense cDNA.
XX
KW Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200179505-A2.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-US12435.
XX
PR 17-APR-2000; 2000US-197977P.
PR 17-APR-2001; 2001US-0836410.
XX
PA (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI Gendron RL, Paradis H;
XX
DR WPI; 2002-017618/02.
XX
PT Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT overexpression of tubedown-1 protein and are useful to treat
PT osteosarcoma and Ewing's Sarcoma family of tumours -
XX
PS Claim 6; Page 38-39; 56pp; English.
XX
CC The sequence represents tubedown-1 (tbdn-1) bases 1413-1 antisense cDNA.
CC The invention relates to a novel isolated nucleic acid of the tubedown-1
CC gene, and antisense nucleic acids to tbdn-1. The polynucleotides and
CC protein of the invention have cytostatic and osteopathic activity. The
CC polynucleotides of the invention may be used in antisense-therapy/gene
CC therapy. They are useful in the treatment of bone tumours, especially
CC osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC invention may also be useful for the prevention of metastases from these
CC types of tumours, either alone or in combination with radiotherapy and/or
CC chemotherapeutic agents.
XX
SQ Sequence 1413 BP; 359 A; 314 C; 244 G; 496 T; 0 other;

Query Match 41.3%; Score 1413; DB 24; Length 1413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 caagtaacacccgcaagatgatagaggatctgcagagtgagcatcatggattggttatgc 60
Db   1413 CAAGTAACACCCGCAAGATGATAGAGGATCTGCAGAGTGAGCATCATGGATTGGTTATGC 1354

Qy     61 tttaccatttattagaagactatgaaatggcagcaaaaattttagaagagtttaggaaaa 120
Db   1353 TTTACCATTTATTAGAAGACTATGAAATGGCAGCAAAAATTTTAGAAGAGTTTAGGAAAA 1294

Qy    121 cacagcagacatctcctgataaagtggattatgaatatagtgaactcctcttatatcaga 180
Db   1293 CACAGCAGACATCTCCTGATAAAGTGGATTATGAATATAGTGAACTCCTCTTATATCAGA 1234
```

```
Qy    181 atcaagttcttcgggaagcaggtctttatagagaagccctggaacatctttgtacctatg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ATCAAGTTCTTCGGGAAGCAGGTCTTTATAGAGAAGCCCTGGAACATCTTTGTACCTATG 1174

Qy    241 aaaagcagatttgtgataaacttgctgttgaagaaaccaaaggggaacttctgttgcagt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 AAAAGCAGATTTGTGATAAACTTGCTGTTGAAGAAACCAAAGGGGAACTTCTGTTGCAGT 1114

Qy    301 tgtgtcgtttggaagatgctgctgacgtttatagaggattacaagagaggaatcctgaaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 TGTGTCGTTTGGAAGATGCTGCTGACGTTTATAGAGGATTACAAGAGAGGAATCCTGAAA 1054

Qy    361 attgggcctattacaaaggcttagaaaaagcactgaagccagctaatatgttagaacggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 ATTGGGCCTATTACAAAGGCTTAGAAAAAGCACTGAAGCCAGCTAATATGTTAGAACGGC 994

Qy    421 taaaaatatatgaggaagcctggactaaataccccaggggactcgtgccaagaaagctgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 TAAAAATATATGAGGAAGCCTGGACTAAATACCCCAGGGGACTCGTGCCAAGAAAGCTGC 934

Qy    481 ccttaaacttttatctggagagaagtttaaggagtgtttggataggttcctaaggatga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 CCTTAAACTTTTTATCTGGAGAGAAGTTTAAGGAGTGTTTGGATAGGTTCCTAAGGATGA 874

Qy    541 atttcagcaagggctgtccacctgtcttcaatacctttgaggtctttatacagagataaag 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ATTTCAGCAAGGGCTGTCCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAG 814

Qy    601 agaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagttgtc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 AGAAGGTGGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTTGTC 754

Qy    661 gcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgggtcc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 GCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACATTACTTTGGGTCC 694

Qy    721 agtactatttggcacagcattatgataaaattggtcagccatccattgctctggaatataca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 AGTACTATTTGGCACAGCATTATGATAAAATTGGTCAGCCATCCATTGCTCTGGAATACA 634

Qy    781 taaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagctaaaa 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 TAAATACTGCAATTGAAAGTACACCAACATTGATAGAACTCTTTCTTGTAAAAGCTAAAA 574

Qy    841 tctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccaggccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 TCTATAAGCATGCTGGGAATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCC 514

Qy    901 tggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagccaacc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 TGGACACAGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCAACC 454

Qy    961 tgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcggtag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 TGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGAACTTCAGCGGTAG 394

Qy   1021 agaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatacaaag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     393 AGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGACAGAGTGTGCTCAGGCATACAAAG 334

Qy   1081 caatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacattttatag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 CAATGAACAAATTTGGTGAAGCACTTAAGAAATGTCATGAAATTGAGAGACATTTTATAG 274

Qy   1141 aaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacccttagat 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 AAATCACCGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGAT 214

Qy   1201 catatgtggacttattaaaactagaagatgtacttcgacagcatccattttacttcaaag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     213 CATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAAG 154

Qy   1261 cagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgacagatgaga 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     153 CAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCTCTGACAGATGAGA 94

Qy   1321 acaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaactgcgta 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93 ACAAAGAACACGAGGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAACTGCGTA 34

Qy   1381 ataaacaaagaagagctcaaaagaaagcccaga 1413
          ||||||||||||||||||||||||||||||||
Db      33 ATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGA 1
```

RESULT 6
AAH16408
ID AAH16408 standard; cDNA; 1802 BP.
XX
AC AAH16408;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15380.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15380; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX
SQ Sequence 1802 BP; 644 A; 298 C; 395 G; 465 T; 0 other;

Query Match 35.4%; Score 1210.2; DB 22; Length 1802;
Best Local Similarity 92.3%; Pred. No. 1.2e-271;
Matches 1287; Conservative 0; Mismatches 103; Indels 5; Gaps 1;

```
Qy     27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81
Db    408 gacctgcgcagagagcatcatggattggttatgctattgcttaccatttattagaagatt 467

Qy     82 atgaaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgata 141
Db    468 atgaaatggcagcaaagattttagaagaatttaggaaaacacaacagacatcccctgaca 527

Qy    142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201
Db    528 aggtggattatgaatatagtgaactactcttatatcagaatcaagttcttcgggaagcag 587

Qy    202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261
Db    588 gtctctatagagaagctttggaacatctttgtacctatgaaaagcagatttgtgataaac 647

Qy    262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321
Db    648 ttgctgtagaagaaaccaaaggggaacttctgttgcaactatgtcgtttggaagatgctg 707

Qy    322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381
Db    708 cagatgtttatagaggattgcaagagagaaatcctgaaaactgggcctattacaaaggct 767

Qy    382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441
Db    768 tggaaaaagcactcaagccagctaatatgttagaacggctaaaaatttatgaggaagcct 827

Qy    442 ggactaaataccccaggggactcgtgccaagaaagctgcccttaaacttttatctggag 501
Db    828 agactaaatatcccaggggactggtgccaagaaggctgccgttaaacttttatctggtg 887

Qy    502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561
Db    888 agaagtttaaagaatgtttggataagttcctaaggatgaatttcagcaagggttgcccac 947

Qy    562 ctgtcttcaatacttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621
Db    948 cagtcttcaatactttaagatcattatacaaagacaaagaaaaggtggcaatcatagaag 1007

Qy    622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681
Db   1008 agttagtagtaggttatgaaacctctctaaaaagctgccggttatttaaccccaatgatg 1067

Qy    682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741
Db   1068 atggaaaggaggaaccaccaaccacattactttgggtccagtactacttggcacaacatt 1127

Qy    742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801
Db   1128 atgacaaaattggtcagccatctattgctttggagtacataaatactgctattgaaagta 1187

Qy    802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861
Db   1188 cacctacattaatagaactctttctcgtgaaagctaaaatctataagcatgctggaaata 1247

Qy    862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921
Db   1248 ttaaagaagctgcaaggtggatggatgaggcccaggccttggacacagcagacagattta 1307

Qy    922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981
Db   1308 tcaactccaaatgtgcaaaatacatgctaaaagccaacctgattaaagaagctgaagaaa 1367

Qy    982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041
Db   1368 tgtgctcaaagtttacaagggaaggaacatcagcggtagagaatttgaatgaaatgcagt 1427

Qy   1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101
Db   1428 gcatgtggttccaaacagaatgtgcccaggcttataaagcaatgaataaatttggtgaag 1487

Qy   1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
Db   1488 cacttaagaaatgtcacgagattgagagacattttatagaaatcactgatgaccagtttg 1547

Qy   1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
Db   1548 actttcatacatactgtatgaggaagattacccttagatcatatgtggacttattaaaac 1607

Qy   1222 tagaagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgaga 1281
Db   1608 tagaagatgtacttcgacagcatccatttttacttcaaggcagcaagaattgctatagaga 1667

Qy   1282 tctatttgaagcttcatgacaaccctctgacagatgagaacaaagaacacgaggctgata 1341
Db   1668 tctatttgaagcttcatgacaaccccctacagatgagaataaagaacacgaagctgata 1727

Qy   1342 cagcaaacatgtctgacaaagagctaaagaaactgcgtaataaacaaagaagagctcaaa 1401
Db   1728 cagcaaacatgtctgacaaagagctaaagaagctacgtaataaacaaagaagagctcaaa 1787

Qy   1402 agaaagcccagattg 1416
Db   1788 agaaagcccagatag 1802
```

RESULT 7
AAH14477
ID AAH14477 standard; cDNA; 2859 BP.
XX
AC AAH14477;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11977.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11977; 2537pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2859 BP; 921 A; 476 C; 547 G; 915 T; 0 other;

Query Match 32.9%; Score 1126; DB 22; Length 2859;
Best Local Similarity 90.2%; Pred. No. 5.6e-252;
Matches 1407; Conservative 0; Mismatches 105; Indels 48; Gaps 17;

```
Qy   1882 gattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaatag 1941
          ||||||||||||||||||||||||||||||| |||||| |||||||||| || ||||| ||||
Db      1 gattatcagctgccaaaatggtatattacttagatccttctagtcagaagcgagctatag 60

Qy   1942 agctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatggaag 2001
          || |||| |||||||||||| ||| ||||| |||||||||| ||||| || ||||| |
Db     61 agttggcaacaacacttgatgaatctctcactaacagaaacctccagacatgtatggagg 120

Qy   2002 tgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcctaca 2061
          | |||||||||||||| |||||||||||||| |||||||||||||||||||| ||| || |
Db     121 tattggaagccttgtatgatggtagcctaggagactgtaaagaagctgctgaaatttata 180

Qy   2062 gagcaagttgtcataagcttttcccttatgctttggctttcatgcctcctggatacgaag 2121
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     181 gagcaaattgtcataagcttttcccttatgctttggctttcatgcctcctggatatgaag 240

Qy   2122 aggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggccaatg 2181
          |||||||||||||||||| || |||||||||||||||||| | ||||||||||||||||
Db     241 aggatatgaagatcacagttaatggagatagttctgcagaagctgaagaactggccaatg 300

Qy   2182 aaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtataat 2241
          |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     301 aaatttgaacatcactaaacaagcaaatggaatgactttggaccatatctagtatataat 360

Qy   2242 atttttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaatgtt 2301
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     361 atttttgtcacgcacctgctgcattgctctaacttacacagaatgagaggagtaaatgtt 420

Qy   2302 cttgccttcaaata--gtcttacgtttttttatcctgctg-aaaactatatataaaatatc 2358
          |||||||||||||  || |||||||||||||||||| |||| ||||||||||||||||
Db     421 cttgccttcaaatagtgttttacgtttttttatcctgctgaaaaagtatatataaaatatc 480

Qy   2359 taacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgagggg 2418
          |||||||||||||| |||||||||||||||||||||||| |||||||||||||||||| ||
Db     481 taacattacaggatagaggttcagtttcttaaaaaatt-aaagctgctaaaattgagtgg 539

Qy   2419 tttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaactaa 2478
          ||      |||||| |  || |||||||||||| ||  |||  ||    |||| ||||||
Db     540 tt------aaaaaagataccttatcctattcctccccacccacccatgtttttaaactaa 593

Qy   2479 tttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatattac 2538
          |||||||||||||||||||| | |||||||||| ||||||||||||||||||||||||
Db     594 tttatataaaatctggaggctgttacagctaacaaagcaggtgtgtggcagaaatattac 653

```

```
Qy   2539 tttaaatttgtcttgtgagattttgctatatctcagacagcataaataaatgctgtttta 2598
          |||||||||||||||||||||||||| |||||||||||||||    ||||||||||||||
Db     654 tttaaatttgtcttgtgagattttactatatctcagacagc----ataaatgctgtttta 709

Qy   2599 gcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgattct 2658
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| |
Db     710 gcactggattctttcactgagcacaaagagttgttggggctttagcatctgactgatttt 769

Qy   2659 gttacggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacagagcc 2718
          |||||||||||| |   |||||||||||||||||||||||||||||| | |||||||
Db     770 gttacggggttgat-tctgaccataggaagtatgcaatgtgaatcactatttacagagaa 828

Qy   2719 gtctacaacacatgcttgacgttgtagagactgggacacatagctaccaagc-ggattaa 2777
          ||||||||| |||||||| |||||||| |||||||| |||| ||||||||   ||||
Db     829 acctacaacagatgcttgatgttgtagaaactgggacatatagataccaagcaaaattat 888

Qy   2778 gtgaaacctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatgatc 2837
          ||||||||| |||||||||| ||||    |||||||||||||||||||||||||||||||
Db     889 aagaaacctataaggtgttcaatacg---cttgtgtttccaaaattcactgtacatgatc 945

Qy   2838 agtttggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctcaa-t 2896
          |||||||||||||||||||||||||||||| |||||||||||||| |||| |||||| |
Db     946 agtttggtgttcttgtaccacagtttttaactgaaggaaccagttgtaacagtctcaatt 1005

Qy   2897 ttaactaaaacttgaagaactaaaataacaatgcaaacctttatcattgttttggccaaa 2956
          |||||||||||||||||||||||| |||||||||||||||     | ||||||||||
Db     1006 ttaactaaaacttgaagaactaaaacaacaatgcaaacctttcagcattgtttggccaaa 1065

Qy   2957 cttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaattatgc 3016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     1066 cttgttaaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaatta-gc 1124

Qy   3017 aagcatgaatttttcacctgagagtgaaaaaagaaaactctaccatggcttgaagttaca 3076
          |||||||||||||||||  |||||||||||| ||| ||||||||||||||||||| |
Db     1125 aagcatgaatttttcacccaagagtgaaaaaaggaaaatctaccatggcttgaagtta-a 1183

Qy   3077 ggagcagaactcctgactaccattctatgactgatgaagagactaata-tctaaaacctc 3135
          |||||||||||||||||||||||||||||||||||||||||| || ||||||||| |  |||||||
Db     1184 agagcagaactcctgactaccattctatgactgatcaaaagactaatagttaaaaacctc 1243

Qy   3136 agcaggccttgttcacgatatgcag---aaaaagtgctgcagtttagatacctctgggaa 3192
          |||||||||||||||||||||||   |||||||||||||||||||||||||| ||||
Db     1244 agcaggccttgttcacgatatgcagaaaaaaaagtgctgcagtttagatacctct-ggaa 1302

Qy   3193 cttttccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaatatattt 3252
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     1303 tttttccacagtgtcacaggtttgtaatacttgaagccctacatttctaagaatatattt 1362

Qy   3253 ctcgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagatttt 3312
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1363 cttgctcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagatttt 1422

Qy   3313 ttttt-------------caataacagaccagcttctttttcctgcagttacaaatgtaa 3359
          |||||                    |||||||||||||||||||||||| |||||||||| ||||||
Db     1423 ttttttttttttttttcaaataacagaccagcttctttttcttgcagttacagatgtaa 1482

Qy   3360 tttctttttttttttgttgtcaaacataaggtaccaaat---atgcaataaattgttttg 3416
          ||||      |||||||||||||||||||||||||||||   ||||||||||||||||
Db     1483 tttc-----ctttttgttgtcaaacataaggtaccaaatatgatgcaataaattgttttg 1537
```

RESULT 8
AAH16424
ID AAH16424 standard; cDNA; 1985 BP.
XX
AC AAH16424;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15407.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15407; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1985 BP; 652 A; 349 C; 492 G; 492 T; 0 other;

Query Match 31.6%; Score 1079.6; DB 22; Length 1985;
Best Local Similarity 91.5%; Pred. No. 3e-241;
Matches 1171; Conservative 0; Mismatches 94; Indels 15; Gaps 2;

```
Qy     27 gatctgcagagtgagcatcatggattggttatgct-----ttaccatttattagaagact 81
          || ||||  || ||||||||||||||||||||||     ||||||||||||||||||| |
Db    710 gacctgcgcagagagcatcatggattggttatgctattgcttaccatttattagaagatt 769

Qy     82 atgaaatggcagcaaaaattttagaagagtttaggaaaacacagcagacatctcctgata 141
          |||||||||||||||| |||||||||| |||||||||||||| |||||||| ||||| |
Db    770 atgaaatggcagcaaagattttagaagaatttaggaaaacacaacagacatcccctgaca 829

Qy    142 aagtggattatgaatatagtgaactcctcttatatcagaatcaagttcttcgggaagcag 201
          | ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    830 aggtggattatgaatatagtgaactactcttatatcagaatcaagttcttcgggaagcag 889

Qy    202 gtctttatagagaagccctggaacatctttgtacctatgaaaagcagatttgtgataaac 261
          |||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    890 gtctctatagagaagctttggaacatctttgtacctatgaaaagcagatttgtgataaac 949

Qy    262 ttgctgttgaagaaaccaaaggggaacttctgttgcagttgtgtcgtttggaagatgctg 321
          ||||||| |||||||||||||||||||||||||||||||  | ||||||||||||||||||
Db    950 ttgctgtagaagaaaccaaaggggaacttctgttgcaactatgtcgtttggaagatgctg 1009

Qy    322 ctgacgtttatagaggattacaagagaggaatcctgaaaattgggcctattacaaaggct 381
          | || |||||||||||||| |||||||| |||||||||| ||||||||||||||||||||
Db   1010 cagatgtttatagaggattgcaagagagaaatcctgaaaactgggcctattacaaaggct 1069

Qy    382 tagaaaaagcactgaagccagctaatatgttagaacggctaaaaatatatgaggaagcct 441
          | |||||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db   1070 tggaaaaagcactcaagccagctaatatgttagaacggctaaaaatttatgaggaagcct 1129

Qy    442 ggactaaatacccccaggggactcgtgccaagaaagctgcccttaaacttttttatctggag 501
          |||||||||| |||||||||| |||||||||| |||||| ||||||||||||||||||| |
Db   1130 ggactaaatatcccaggggactggtgccaagaaggctgccgttaaacttttttatctggtg 1189

Qy    502 agaagtttaaggagtgtttggataggttcctaaggatgaatttcagcaagggctgtccac 561
          |||||||||| || |||||||||| |||||||||||||||||||||||||||||| || ||||
Db   1190 agaagtttaaagaatgtttggataagttcctaaggatgaatttcagcaagggttgcccac 1249

Qy    562 ctgtcttcaatacctttgaggtctttatacagagataaagagaaggtggcaatcgtagaag 621
          | |||||||||||| || || || ||||||| ||| ||||| |||||||||||| ||||||
Db   1250 cagtcttcaatactttaagatcattatacaaagacaaagaaaaggtggcaatcatagaag 1309

Qy    622 aactagtagttggttatgaaacttctctaaaaagttgtcgcctatttaaccccaatgatg 681
          |  ||||||| |||||||||| |||||||||| || ||  ||||||||||||||||||||
Db   1310 agttagtagtaggttatgaaacctctctaaaaagctgccggttatttaaccccaatgatg 1369

Qy    682 atggaaaggaggaacctccaaccacattactttgggtccagtactatttggcacagcatt 741
          |||||||||||||||| |||||||||||||||||||||||||||||| |||||||| ||||
Db   1370 atggaaaggaggaaccaccaaccacattactttgggtccagtactacttggcacaacatt 1429

Qy    742 atgataaaattggtcagccatccattgctctggaatacataaatactgcaattgaaagta 801
          |||| |||||||||||||||| |||||| |||| |||||||||||| ||||||||||
Db   1430 atgacaaaattggtcagccatctattgctttggagtacataaatactgctattgaaagta 1489

Qy    802 caccaacattgatagaactctttcttgtaaaagctaaaatctataagcatgctgggaata 861
          |||| ||||| |||||||||||||| || |||||||||||||||||||||||||| ||||
Db   1490 cacctacattaatagaactctttctcgtgaaagctaaaatctataagcatgctggaaata 1549

Qy    862 ttaaagaagctgccaggtggatggatgaagcccaggccctggacacagcagacagattta 921
          |||||||||||||| |||||||||||||| ||||||||| ||||||||||||||||||||
Db   1550 ttaaagaagctgcaaggtggatggatgaggcccaggccttggacacagcagacagattta 1609

Qy    922 ttaattccaagtgtgcaaaatacatgttaaaagccaacctgattaaagaggctgaagaaa 981
          | || ||||| |||||||||||||||| ||||||||||||||||||||||| ||||||||
Db   1610 tcaactccaaatgtgcaaaatacatgctaaaagccaacctgattaaagaagctgaagaaa 1669

Qy    982 tgtgttccaagtttacgagggaaggaacttcagcggtagagaacctgaatgaaatgcagt 1041
          |||| || |||||||| |||||||||| |||||||||||||  ||||||||||||||||
Db   1670 tgtgctcaaagtttacaagggaaggaacatcagcggtagagaatttgaatgaaatgcagt 1729

Qy   1042 gtatgtggttccagacagagtgtgctcaggcatacaaagcaatgaacaaatttggtgaag 1101
          | |||||||||| ||||| ||||| ||||| || |||||||||| ||||||||||||||
Db   1730 gcatgtggttccaaacagaatgtgcccaggcttataaagcaatgaataaatttggtgaag 1789

Qy   1102 cacttaagaaatgtcatgaaattgagagacattttatagaaatcaccgatgaccagtttg 1161
          |||||||||||||||||| |||||||||          ||||||| ||||||||||||||
Db   1790 cacttaagaaatgtcatgagattgagaga----------aaatcactgatgaccagtttg 1839

Qy   1162 actttcatacatactgtatgaggaagatcacccttagatcatatgtggacttattaaaac 1221
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1840 actttcatacatactgtatgaggaagattacccttagatcatatgtggacttattaaaac 1899

Qy   1222 tagaagatgtacttcgacagcatccatttttacttcaaagcagcgagaattgctattgaga 1281
          |||||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||||| ||||
Db   1900 tagaagatgtacttcgacagcatccatttttacttcaaggcagcaagaattgctatagaga 1959

Qy   1282 tctatttgaagcttcatgac 1301
          ||||||||||||||||||||
Db   1960 tctatttgaagcttcatgac 1979
```

```
RESULT   9
AAH81664/c
ID   AAH81664 standard; DNA; 710 BP.
XX
AC   AAH81664;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human differential transcription-associated cDNA SEQ ID 173.
XX
KW   Differential transcription; human; rat; tumour cell; cytostatic;
KW   Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157058-A2.
XX
PD   09-AUG-2001.
XX
PF   31-JAN-2001; 2001WO-EP01003.
XX
PR   31-JAN-2000; 2000DE-1004102.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Hinzmann B,  Schaefer R,  Zuber J,  Tchernitsa O;
PI   Grips M,  Hellriegel M,  Schmitz A,  Sers C;
XX
DR   WPI; 2001-483415/52.
XX
PT   Nucleic acids differentially expressed between tumor and normal cells,
PT   useful for diagnosis or therapy of tumors and for screening active
PT   agents -
XX
PS   Disclosure; Page 376; 579pp; German.
XX
CC   This invention describes a nucleic acid (I) with differential expression
CC   between tumour and normal cells and which has cytostatic activity. (I)
CC   work as modulators of Ras activity by inducing expression of tumour
CC   suppressor genes. (I), and polypeptides encoded by them, are useful as
CC   targets for diagnosis or therapy and in screening to determine the
CC   effects of an active compound (potential pharmaceutical) on a cell line,
CC   particularly for diagnosis and treatment of tumors, especially by
CC   modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC   methods) or by modulating the amount and/or location of (I)-encoded
CC   polypeptides (by administration of the polypeptide or its activator,
CC   antibody (optionally as a conjugate) or inhibitor). The method allows
CC   identification of many Class II tumour suppressor genes (i.e. genes that
CC   are not primary targets for tumour-initiating mutations).
CC   AAH81492-AAH82376 represent the human and rat derived nucleic acid
CC   fragments described in the method of the invention.
XX
SQ   Sequence 710 BP; 170 A; 146 C; 121 G; 267 T; 6 other;

  Query Match             18.1%;  Score 617.8;  DB 22;  Length 710;
  Best Local Similarity   92.4%;  Pred. No. 5.8e-134;
  Matches  657;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy    714 tgggtccagtactatttggcacagcattatgataaaattggtcagccatccattgctctg 773
          |||||| ||||||| |||||||| |||||||| |||||||||||||||| | |||| ||
Db    710 TGGGTCAAGTACTACTTGGCACACCATTATGACAAAATTGGTCAGCCATCTANTGCTNTG 651

Qy    774 gaatacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaa 833
          || |||||||||||||| ||||| |||||| | ||||| |||||  ||||||| || | |
Db    650 GAGTACATAAATACTGCTATTGANAGTACA-CTACATTAATAGACNTCTTTCTCGTGANA 592

Qy    834 gctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcc 893
          |||||||||||||||||||||| |||||||||||||||| |||||||||||||| ||
Db    591 GCTAAAATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCN 532

Qy    894 caggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaa 953
          |||||| |||||||||||||||||||||| || ||||| ||||||||||||||| |||||
Db    531 CAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAA 472

Qy    954 gccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttca 1013
          |||||||||||||||| |||||||||||||| || |||||||| |||||||||||| |||
Db    471 GCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGAACATCA 412

Qy   1014 gcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggca 1073
          |||||||||||  |||||||||||||||| |||||||||| ||||| ||||| |||||
Db    411 GCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAACAGAATGTGCCCAGGCT 352

Qy   1074 tacaaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacat 1133
          || |||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||
Db    351 TATAAAGCAATGAATAAATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACAT 292

Qy   1134 tttatagaaatcaccgatgaccagtttgactttcatacatactgtatgaggaagatcacc 1193
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||| |||
Db    291 TTTATAGAAATCACTGATGACCAGTTTGACTTTCATACATACTGTATGAGGAAGATTACC 232

Qy   1194 cttagatcatatgtggacttattaaaactagaagatgtacttcgacagcatccatttttac 1253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CTTAGATCATATGTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTAC 172

Qy   1254 ttcaaagcagcgagaattgctattgagatctatttgaagcttcatgacaaccctctgaca 1313
          ||||| ||||| |||||||||| |||||||||||||||||||||||||| || |||
Db    171 TTCAAGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCCCTTACA 112

Qy   1314 gatgagaacaaagaacacgaggctgatacagcaaacatgtctgacaaagagctaaagaaa 1373
          |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    111 GATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGACAAAGAGCTAAAGAAG 52

Qy   1374 ctgcgtaataaacaaagaagagctcaaaagaaagcccagattgaagaagag 1424
          || |||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     51 CTACGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGATAGAAGAAGAG 1

RESULT  10
AAH06489
ID   AAH06489 standard; cDNA; 790 BP.
XX
AC   AAH06489;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3324.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;  99JP-0248036.
PR   27-AUG-1999;  99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
```

```
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 3324; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 790 BP; 250 A; 136 C; 155 G; 246 T; 3 other;

  Query Match             18.0%;  Score 616.2;  DB 22;  Length 790;
  Best Local Similarity   90.0%;  Pred. No. 1.4e-133;
  Matches  718;  Conservative    0;  Mismatches   66;  Indels   14;  Gaps    5;

Qy   1882 gattatcagctgccaaaatggtatattatttagattcttctagtcaaaaacgagcaatag 1941
           ||||||||||||||||||||||||||||| |||||| |||||||||| || ||||| ||||
Db      1 gattatcagctgccaaaatggtatattacttagatccttctagtcagaagcgagctatag 60

Qy   1942 agctggcgacaacacttgatggatccctcaccaacagaaaccttcagacttgcatggaag 2001
           || |||| |||||||||||||| ||| ||||| |||||||||||| ||||| || ||||| |
Db     61 agttggcaacaacacttgatgaatctctcactaacagaaacctccagacatgtatggagg 120

Qy   2002 tgttggaagccttgtgtgatggtagcctacgagactgtaaagaagctgccgaagcctaca 2061
           | |||||||||||||| |||||||||||||| ||||||||||||||||||| |||   || |
Db    121 tattggaagccttgtatgatggtagcctaggagactgtaaagaagctgctgaaatttata 180

Qy   2062 gagcaagttgtcataagcttttcccttatgctttggctttcatgcctcctggatacgaag 2121
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    181 gagcaaattgtcataagcttttcccttatgctttggctttcatgcctcctggatatgaag 240

Qy   2122 aggatatgaagatcacagtgaacggagatagttctgcagaaacggaagaactggccaatg 2181
           |||||||||||||||||| || |||||||||||||||||| | ||||||||||||||||
Db    241 aggatatgaagatcacagttaatggagatagttctgcagaagctgaagaactggccaatg 300

Qy   2182 aaatctgaacatcattaaacaagcaaatggaatgactttggaccatatctagtgtataat 2241
           |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||| ||||||
Db    301 aaatttgaacatcactaaacaagcaaatggaatgactttggaccatatctagtatataat 360

Qy   2242 atttttgtcacgcacctgctgcattgctcttacttacacagaatgagaggagtaaatgtt 2301
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    361 atttttgtcacgcacctgctgcattgctctaacttacacagaatgagaggagtaaatgtt 420

Qy   2302 cttgccttcaaatagtc--ttacgtttttttatcctgctg-aaaactatatataaaatatc 2358
           ||||||||||||||||   |||||||||||||||||||| |||| ||||||||||||||||
Db    421 cttgccttcaaatagtgttttacgtttttttatcctgctgaaaaagtatatataaaatatc 480

Qy   2359 taacattacaggatataggttcagtttcttaaaaaattaaaagctgctaaaattgagggg 2418
           |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||| ||
Db    481 taacattacaggatagaggttcagtttcttaaaaaatt-aaagctgctaaaattgagtgg 539

Qy   2419 tttaaaagaaaaaaaaatccgtatcctattcctaccttcccttcccatgtttttaaactaa 2478
           ||      |||||| | || |||||||||||| || ||| ||    |||| ||||||
Db    540 tt------aaaaaagataccttatcctattcctccccacccacccatgtttttaaactaa 593
```

```
Qy   2479 tttatataaaatctggaggctataacagctaacatagcaggtgtgtggcagaaatattac 2538
           |||||||||||||||||||||| | |||||||||| |||||||||||||||||||||||||
Db    594 tttatataaaatctggaggctgttacagctaacaaagcaggtgtgtggcagaaatattac 653

Qy   2539 tttaaatttgtcttgtgagattttgctatatctcagacagcataaataaatgctgtttta 2598
           |||||||||||||||||||||||| |||||||||||||||    |||||||||| ||||
Db    654 tttaaatttgtcttgtgagattttactatatctcagacagc----ataaatgctggttta 709

Qy   2599 gcactggattctttcactgagcacaaagagttgttggggctttagcatctgcctgattct 2658
           |||||||||||||||||||||||||||||||||| ||||||||||||||| |||  | |
Db    710 gcactggattctttcactgagcacaaagagttgntggggctttagcatctgactggattt 769

Qy   2659 gttacggggttggtgatt 2676
           | ||| ||| || |  ||
Db    770 ggtacngggntgattctt 787
```

```
RESULT  11
AAX99092
ID   AAX99092 standard; cDNA; 1225 BP.
XX
AC   AAX99092;
XX
DT   24-SEP-1999  (first entry)
XX
DE   Human cancer cell derived cDNA contig #18.
XX
KW   Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW   integral membrane protein; aspartyl protease; GATA family; wnt family;
KW   transcription factor; G-protein alpha subunit; protein phosphatase;
KW   phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW   protein kinase; tyrosine phosphatase; developmental signalling protein;
KW   WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW   detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW   Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW   leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW   prostate; ss.
XX
OS   Homo sapiens.
XX
PN   WO9933982-A2.
XX
PD   08-JUL-1999.
XX
PF   22-DEC-1998;   98WO-US27610.
XX
PR   21-DEC-1998;   98US-0217471.
PR   23-DEC-1997;   97US-0068755.
PR   03-APR-1998;   98US-0080664.
PR   21-OCT-1998;   98US-0105234.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones LW,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-430243/36.
XX
PT   New isolated human polynucleotides
XX
PS   Claim 1; Page 580; 591pp; English.
XX
CC   This invention describes novel isolated human polynucleotides obtained
CC   by screening for differential expression in colon cancer, breast cancer
CC   and lung cancer cell lines. The polynucleotides of the invention are
CC   represented in AAX98275-X99118 and encode polypeptides of protein
CC   families selected from 4 transmembrane segments integral membrane
```

```
CC   proteins, 7 transmembrane receptors, ATPases associated with various
CC   cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC   transcription factors, G-protein alpha subunit, phorbolesters or
CC   diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC   protein tyrosine phosphatase, trypsin, wnt family of developmental
CC   signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC   encoded polypeptides also have a functional domain selected from Ank
CC   repeat, basic region plus leucine zipper transcription factors,
CC   bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC   (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC   domain. The polynucleotides encode polypeptides with similarity to known
CC   protein families and are predicted to have similar properties. The novel
CC   polynucleotides can be used to develop products for use as therapeutic
CC   agents and in forensics, genetic analysis, mapping and diagnostic
CC   applications. In particular, the product can be used for the detection
CC   and management of cancers. They can be used for treating e.g. cervical
CC   cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC   retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC   myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC   myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC   hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC   epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC   mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC   prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC   the skin.
XX
SQ   Sequence 1225 BP; 400 A; 203 C; 240 G; 379 T; 3 other;

  Query Match             16.5%;  Score 565;  DB 20;  Length 1225;
  Best Local Similarity   89.7%;  Pred. No. 1.3e-121;
  Matches  749;  Conservative    1;  Mismatches   51;  Indels   34;  Gaps   12;

Qy   2604 ggattctttcactgagcacaaagagttgttggggctttagcatctgcctgattctgttac 2663
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| ||||||
Db      1 ggattctttcactgagcacaaagagttgttggggctttagcatctgactgattttgttac 60

Qy   2664 ggggttggtgattgaccataggaagtatgcaatgtgaatcactgtgtacagagccgtcta 2723
           ||||||| |   |||||||||||||||||||||||||||||| | |||||||    ||| 
Db     61 ggggttgat-tctgaccataggaagtatgcaatgtgaatcactatttacagagaaacctа 119

Qy   2724 caacacatgcttgacgttgtagagactgggacacatagctaccaagc-ggattaagtgaa 2782
           ||||| |||||||| |||||||| ||||||||| |||| ||||||||   ||||   ||| 
Db    120 caacagatgcttgatgttgtagaaactgggacatatagataccaagcaaaattataagaa 179

Qy   2783 acctagaaggtgttcagtacgtgtgttgtgtttccaaaattcactgtacatgatcagttt 2842
           ||||| |||||||||| ||||    |||||||||||||||||||||||||||||||||| 
Db    180 acctataaggtgttcaatacg---cttgtgtttccaaaattcactgtacatgatcagttt 236

Qy   2843 ggtgttcttgtaccacagtttttaaccgaaggaaccagttggaacaatctcaa-tttaac 2901
           |||||||||||||||||||||||| |||||||||||||| |||| |||||| |||||| 
Db    237 ggtgttcttgtaccacagtttttaactgaaggaaccagttgtaacagtctcaattttaac 296

Qy   2902 taaaacttgaagaactaaaataacaatgcaaacctttatcattgttttggccaaacttgt 2961
           |||||||||||||||||| ||||||||||||||||    |  ||||||||||||||| 
Db    297 taaaacttgaagaactaaaacaacaatgcaaacctttcagcattgtttggccaaacttgt 356

Qy   2962 taaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaattatgcaagca 3021
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| 
Db    357 taaaactgtaatgcaagaaccaaatgcactgtgatgtggcaccaactaatta-gcaagca 415

Qy   3022 tgaatttttcacctgagagtgaaaaaagaaaactctaccatggcttgaagttacaggagc 3081
           |||||||||||||  |||||||||||||| ||| ||||||||||||||||||| | |||| 
Db    416 tgaatttttcacccaagagtgaaaaaaggaaaatctaccatggcttgaagtta-aagagc 474

Qy   3082 agaactcctgactaccattctatgactgatgaagagactaata-tctaaaacctcagcag 3140
           |||||||||||||||||||||||||||||| || |||||||| |  |||||||||||| 
Db    475 agaactcctgactaccattctatgactgatcaaaagactaatagttaaaaacctcagcag 534

Qy   3141 gccttgttcacgatatgcag---aaaaagtgctgcagtttagatacctctgggaacttttt 3197
           ||||||||||||||||||   |||||||||||||||||||||||||| |||| |||| 
Db    535 gccttgttcacgatatgcagaaaaaaaagtgctgcagtttagatacctct-ggaattttt 593

Qy   3198 ccacagtgtcacaggtttgtaatacttgaagcccttcatttctaagaatatatttctcgc 3257
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||| || 
Db    594 ccacagtgtcacaggtttgtaatacttgaagccctacatttctaagaatatatttcttgc 653

Qy   3258 tcagttgtttcaggcaagcccaagactttgtaatttttaaagggcccaagatttttttttt 3317
           |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    654 tcagttgtttcakgcaagcccaagactttgtaatttttaaagggcccaagatttttttttt 713

Qy   3318 -------------caataacagaccagcttctttttttcctgcagttacaaatgtaatttct 3364
                        |||||||||||||||||||||||||||| |||||||||| |||||||||| 
Db    714 ttttttttttttcaaataacagaccagcttctttttcttgcagttacagatgtaatttc- 772

Qy   3365 ttttttttttgttgtcaaacataaggtaccaaat---atgcaataaattgttttg 3416
                 ||||||||||||||||||||||||||||||   |||||||||||||||||| 
Db    773 ----ctttttgttgtcaaacataaggtaccaaatatgatgcaataaattgttttg 823


RESULT  12
AAZ15705
ID   AAZ15705 standard; cDNA; 781 BP.
XX
AC   AAZ15705;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:3174.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1524-1525; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
```

CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 781 BP; 263 A; 140 C; 156 G; 205 T; 17 other;

Query Match 16.3%; Score 555.6; DB 20; Length 781;
Best Local Similarity 90.8%; Pred. No. 1.8e-119;
Matches 621; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```
Qy    597 aaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagt 656
          ||  | ||||||||||||| |||||||  ||||||| |||||||||||| ||||||||||
Db     94 aagaaaaaggtggcaatcatagaagagttagtagtaggttatgaaacctctctaaaaagc 153

Qy    657 tgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccacattactttgg 716
          || ||  ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    154 tgccggttatttaaccccaatgatgatggaaaggaggaaccaccaaccacattactttgg 213

Qy    717 gtccagtactatttggcacagcattatgataaaattggtcagccatccattgctctggaa 776
          ||||  | ||| |||||||| |||||||| |||||||||||||||||| |||||| ||||
Db    214 gtccnntnctacttggcacaacattatgacaaaattggtcagccatctattgctttggag 273

Qy    777 tacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagct 836
          |||||||||||||| |||||||||||||| ||||| |||||||||||||| || ||||||
Db    274 tacataaatactgctattgaaagtacacctacattaatagaactctttctcgtgaaagct 333

Qy    837 aaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccag 896
          ||||||||||||||||||| |||||||||||||||||| |||||||||||||| ||||||
Db    334 aaaatctataagcatgctggaaatattaaagaagctgcaaggtggatggatgaggcccag 393

Qy    897 gccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagcc 956
          ||| |||||||||||||||||||||| || ||||| |||||||||||||| ||||||||
Db    394 gccttggacacagcagacagatttatcaactccaaatgtgcaaaatacatgctaaaagcc 453

Qy    957 aacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcg 1016
          |||||||||||||| |||||||||||||| || |||||||| |||||||||||| ||||||
Db    454 aacctgattaaagaagctgaagaaatgtgctcaaagtttacaagggaaggaacatcagcg 513

Qy   1017 gtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatac 1076
          ||||||||  |||||||||||||||| |||||||||||| ||||| ||||| ||||| ||
Db    514 gtagagaatttgaatgaaatgcagtgcatgtggttccaaacagaatgtgcccaggcttat 573

Qy   1077 aaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacatttt 1136
          |||||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||
Db    574 aaagcaatgaataaatttggtgaagcacttaagaaatgtcatgagattgagagacatttt 633

Qy   1137 ata-gaaatcaccgatgaccagtttgactttcatacatactgtatg-aggaagatcaccc 1194
          ||| |||||||| |||||||||||||||||||||||||||||||| ||| |||||||| ||||
Db    634 ataggaaatcactgatgaccagtttgactttcatacatactggatgaaggaagattaccc 693

Qy   1195 ttagatcatatgtggacttattaaaacta-gaagatgtacttcgacagcatccattttac 1253
          |||||||||||||||||||||| |||||| |||||||||||  ||||||| ||||||||
Db    694 ttagatcatatgtggacttattnaaactatgaagatgtactttnacagcatncattttac 753

Qy   1254 ttcaaagcagcgagaattgctatt 1277
          ||||| ||||| ||||||||| ||
Db    754 ttcaaggcagcaagaattgctttt 777
```

RESULT 13
AAX99053
ID AAX99053 standard; cDNA; 781 BP.
XX
AC AAX99053;
XX
DT 24-SEP-1999 (first entry)

XX
DE Human validated cancer cell derived cDNA #375.
XX
KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW integral membrane protein; aspartyl protease; GATA family; wnt family;
KW transcription factor; G-protein alpha subunit; protein phosphatase;
KW phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW protein kinase; tyrosine phosphatase; developmental signalling protein;
KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW prostate; ss.
XX
OS Homo sapiens.
XX
PN WO9933982-A2.
XX
PD 08-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US27610.
XX
PR 21-DEC-1998; 98US-0217471.
PR 23-DEC-1997; 97US-0068755.
PR 03-APR-1998; 98US-0080664.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-430243/36.
XX
PT New isolated human polynucleotides
XX
PS Claim 1; Page 564; 591pp; English.
XX
CC This invention describes novel isolated human polynucleotides obtained
CC by screening for differential expression in colon cancer, breast cancer
CC and lung cancer cell lines. The polynucleotides of the invention are
CC represented in AAX98275-X99118 and encode polypeptides of protein
CC families selected from 4 transmembrane segments integral membrane
CC proteins, 7 transmembrane receptors, ATPases associated with various
CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC transcription factors, G-protein alpha subunit, phorbolesters or
CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC protein tyrosine phosphatase, trypsin, wnt family of developmental
CC signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC encoded polypeptides also have a functional domain selected from Ank
CC repeat, basic region plus leucine zipper transcription factors,
CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC domain. The polynucleotides encode polypeptides with similarity to known
CC protein families and are predicted to have similar properties. The novel
CC polynucleotides can be used to develop products for use as therapeutic
CC agents and in forensics, genetic analysis, mapping and diagnostic
CC applications. In particular, the product can be used for the detection
CC and management of cancers. They can be used for treating e.g. cervical
CC cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC the skin.

XX
SQ Sequence 781 BP; 263 A; 140 C; 156 G; 205 T; 17 other;

Query Match 16.3%; Score 555.6; DB 20; Length 781;
Best Local Similarity 90.8%; Pred. No. 1.8e-119;
Matches 621; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```
Qy    597 aaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttctctaaaaagt 656
          ||  | ||||||||||||  |||||||  |||||||  ||||||||||| ||||||||||
Db     94 aagaaaaaggtggcaatcatagaagagttagtagtaggttatgaaacctctctaaaaagc 153

Qy    657 tgtcgcctatttaaccccaatgatgatggaaaggaggaaccttccaaccacattactttgg 716
          || ||  ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    154 tgccggttatttaaccccaatgatgatggaaaggaggaaccaccaaccacattactttgg 213

Qy    717 gtccagtactatttggcacagcattatgataaaattggtcagccatccattgctctggaa 776
          ||||  | |||  |||||||| |||||||| |||||||||||||||| |||||| ||||
Db    214 gtccnntnctacttggcacaacattatgacaaaattggtcagccatctattgctttggag 273

Qy    777 tacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgtaaaagct 836
          |||||||||||||| |||||||||||||| ||||| |||||||||||||| || ||||||
Db    274 tacataaatactgctattgaaagtacacctacattaatagaactctttctcgtgaaagct 333

Qy    837 aaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaagcccag 896
          |||||||||||||||||| |||||||||||||||| |||||||||||||| ||||||
Db    334 aaaatctataagcatgctggaaatattaaagaagctgcaaggtggatggatgaggcccag 393

Qy    897 gccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgttaaaagcc 956
          ||| |||||||||||||||||||||| || ||||| |||||||||||||| ||||||||
Db    394 gccttggacacagcagacagatttatcaactccaaatgtgcaaaatacatgctaaaagcc 453

Qy    957 aacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaacttcagcg 1016
          |||||||||||||| |||||||||||||| || |||||||| |||||||||| ||||||
Db    454 aacctgattaaagaagctgaagaaatgtgctcaaagtttacaagggaaggaacatcagcg 513

Qy   1017 gtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcaggcatac 1076
          ||||||||  |||||||||||||||| |||||||||| ||||| ||||| ||||| ||
Db    514 gtagagaatttgaatgaaatgcagtgcatgtggttccaaacagaatgtgcccaggcttat 573

Qy   1077 aaagcaatgaacaaatttggtgaagcacttaagaaatgtcatgaaattgagagacatttt 1136
          |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||
Db    574 aaagcaatgaataaatttggtgaagcacttaagaaatgtcatgagattgagagacatttt 633

Qy   1137 ata-gaaatcaccgatgaccagtttgactttcatacatactgtatg-aggaagatcaccc 1194
          ||| |||||||| |||||||||||||||||||||||||||||| ||| |||||||| ||||
Db    634 ataggaaatcactgatgaccagtttgactttcatacatactggatgaaggaagattaccc 693

Qy   1195 ttagatcatatgtggacttattaaaacta-gaagatgtacttcgacagcatccatttttac 1253
          |||||||||||||||||||| |||||| |||||||||||  ||||||| ||||||||
Db    694 ttagatcatatgtggacttattnaaactatgaagatgtactttnacagcatncattttac 753

Qy   1254 ttcaaagcagcgagaattgctatt 1277
          ||||| ||||| |||||||||  ||
Db    754 ttcaaggcagcaagaattgctttt 777
```

RESULT 14
AAZ15983
ID AAZ15983 standard; cDNA; 764 BP.
XX
AC AAZ15983;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:3452.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 1650-1651; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 764 BP; 253 A; 138 C; 148 G; 204 T; 21 other;

Query Match 13.9%; Score 475.6; DB 20; Length 764;
Best Local Similarity 90.2%; Pred. No. 7.2e-101;
Matches 552; Conservative 0; Mismatches 56; Indels 4; Gaps 4;

```
Qy    591 agagataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttctcta 650
          ||||| ||||| ||||||||||||| |||||||  | |||| ||||||||||| ||||||
Db     65 agagacaaagaaaaggtggcaatcatagaagagtttntagtaggttatgaaacctctcta 124

Qy    651 aaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccacatta 710
          ||||| || ||  ||||||||||||||||||||||||||||||||||| ||||||||||
Db    125 aaaagctgccggttatttaaccccaatgatgatggaaaggaggaaccaccaaccacatta 184

Qy    711 ctttgggtccagtactatttggcacagcattatgataaaattggtcagccatccattgct 770
          |||||||||||||||| |||||||| |||||||| |||||||||||||||||| ||||||
Db    185 ctttgggtccagtactacttggcacaacattatgacaaaattggtcagccatctattgct 244

Qy    771 ctggaatacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgta 830
          |||| |||||||||||||| |||||||||||||| ||||| |||||||||||||| ||
Db    245 ttggagtacataaatactgctattgaaagtacacctacattaatagaactctttctcgtg 304
```

```
Qy    831 aaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaa 890
          ||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db    305 aaagctaaaatctataagcatgctggaaatattaaagaagctgcaaggtggatggatgag 364

Qy    891 gcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgtta 950
          ||||||||| |||||||||||||||||||| || ||||| ||||||||||||||| ||
Db    365 gcccaggccttggacacagcagacagatttatcaactccaaatgtgcaaaatacatgcta 424

Qy    951 aaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaact 1010
          |||||||||||||||||||| |||||||||||||| || |||||||| ||||||||||
Db    425 aaagccaacctgattaaagaagctgaagaaatgtgctcaaagtttacaagggaaggaaca 484

Qy   1011 tcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcag 1070
          |||||||||||||||  |||||||||||||||| |||||||||| ||||| ||||| |||
Db    485 tcagcggtagagaatttgaatgaaatgcagtgcatgtggttccaaacagaatgtgcccag 544

Qy   1071 gcatacaaagcaatgaa-caaatttggtgaagcacttaagaaatgtca-tgaaattgaga 1128
          || || ||||||||||  |||||||||||||||||||||||||||| ||| |||||||
Db    545 gcttataaagcaatgaattaaatttggtgaagcacttaagaaatgtcattgagattgaga 604

Qy   1129 gacattttatagaaatcaccgatga-ccagtttgactttcatacatactgtatga-ggaa 1186
          ||| |||    |||||||| ||||| ||||||||||||||||||||||||||| ||||
Db    605 gacttttataggaaatcactgatgacccagtttgactttcatacatactgtatganggaa 664

Qy   1187 gatcacccttag 1198
            || ||||||||
Db    665 nattacccttag 676
```

RESULT 15
AAX98777
ID AAX98777 standard; cDNA; 764 BP.
XX
AC AAX98777;
XX
DT 24-SEP-1999 (first entry)
XX
DE Human validated cancer cell derived cDNA #99.
XX
KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW integral membrane protein; aspartyl protease; GATA family; wnt family;
KW transcription factor; G-protein alpha subunit; protein phosphatase;
KW phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW protein kinase; tyrosine phosphatase; developmental signalling protein;
KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW prostate; ss.
XX
OS Homo sapiens.
XX
PN WO9933982-A2.
XX
PD 08-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US27610.
XX
PR 21-DEC-1998; 98US-0217471.
PR 23-DEC-1997; 97US-0068755.
PR 03-APR-1998; 98US-0080664.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-430243/36.
XX
PT New isolated human polynucleotides
XX
PS Claim 1; Page 453; 591pp; English.
XX
CC This invention describes novel isolated human polynucleotides obtained
CC by screening for differential expression in colon cancer, breast cancer
CC and lung cancer cell lines. The polynucleotides of the invention are
CC represented in AAX98275-X99118 and encode polypeptides of protein
CC families selected from 4 transmembrane segments integral membrane
CC proteins, 7 transmembrane receptors, ATPases associated with various
CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC transcription factors, G-protein alpha subunit, phorbolesters or
CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC protein tyrosine phosphatase, trypsin, wnt family of developmental
CC signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC encoded polypeptides also have a functional domain selected from Ank
CC repeat, basic region plus leucine zipper transcription factors,
CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC domain. The polynucleotides encode polypeptides with similarity to known
CC protein families and are predicted to have similar properties. The novel
CC polynucleotides can be used to develop products for use as therapeutic
CC agents and in forensics, genetic analysis, mapping and diagnostic
CC applications. In particular, the product can be used for the detection
CC and management of cancers. They can be used for treating e.g. cervical
CC cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC the skin.
XX
SQ Sequence 764 BP; 253 A; 138 C; 148 G; 204 T; 21 other;

Query Match 13.9%; Score 475.6; DB 20; Length 764;
Best Local Similarity 90.2%; Pred. No. 7.2e-101;
Matches 552; Conservative 0; Mismatches 56; Indels 4; Gaps 4;

```
Qy    591 agagataaagagaaggtggcaatcgtagaagaactagtagttggttatgaaacttctcta 650
          ||||| ||||| |||||||||||| ||||||| | |||| |||||||||||| ||||||
Db     65 agagacaaagaaaaggtggcaatcatagaagagtttntagtaggttatgaaacctctcta 124

Qy    651 aaaagttgtcgcctatttaaccccaatgatgatggaaaggaggaacctccaaccacatta 710
          ||||| || ||  |||||||||||||||||||||||||||||||||||| ||||||||||
Db    125 aaaagctgccggttatttaaccccaatgatgatggaaaggaggaaccaccaaccacatta 184

Qy    711 ctttgggtccagtactatttggcacagcattatgataaaattggtcagccatccattgct 770
          |||||||||||||||| |||||||| |||||||| |||||||||||||||| ||||||
Db    185 ctttgggtccagtactacttggcacaacattatgacaaaattggtcagccatctattgct 244

Qy    771 ctggaatacataaatactgcaattgaaagtacaccaacattgatagaactctttcttgta 830
          |||| |||||||||||||| |||||||||||||| ||||| |||||||||||||| ||
Db    245 ttggagtacataaatactgctattgaaagtacacctacattaatagaactctttctcgtg 304

Qy    831 aaagctaaaatctataagcatgctgggaatattaaagaagctgccaggtggatggatgaa 890
          ||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db    305 aaagctaaaatctataagcatgctggaaatattaaagaagctgcaaggtggatggatgag 364

Qy    891 gcccaggccctggacacagcagacagatttattaattccaagtgtgcaaaatacatgtta 950
          ||||||||| |||||||||||||||||||| || ||||| ||||||||||||||| ||
Db    365 gcccaggccttggacacagcagacagatttatcaactccaaatgtgcaaaatacatgcta 424

Qy    951 aaagccaacctgattaaagaggctgaagaaatgtgttccaagtttacgagggaaggaact 1010
          |||||||||||||||||||| |||||||||||||| || |||||||| ||||||||||
Db    425 aaagccaacctgattaaagaagctgaagaaatgtgctcaaagtttacaagggaaggaaca 484

Qy   1011 tcagcggtagagaacctgaatgaaatgcagtgtatgtggttccagacagagtgtgctcag 1070
```

```
          ||||||||||||||  ||||||||||||||||| |||||||||| ||||| ||||| |||
Db    485 tcagcggtagagaatttgaatgaaatgcagtgcatgtggttccaaacagaatgtgcccag 544

Qy   1071 gcatacaaagcaatgaa-caaatttggtgaagcacttaagaaatgtca-tgaaattgaga 1128
          || || ||||||||||  ||||||||||||||||||||||||||||| ||| |||||||
Db    545 gcttataaagcaatgaattaaatttggtgaagcacttaagaaatgtcattgagattgaga 604

Qy   1129 gacattttatagaaatcaccgatga-ccagtttgactttcatacatactgtatga-ggaa 1186
          ||| |||    |||||||| ||||| ||||||||||||||||||||||||||||| ||||
Db    605 gacttttataggaaatcactgatgacccagtttgactttcatacatactgtatganggaa 664

Qy   1187 gatcacccttag 1198
            || ||||||||
Db    665 nattacccttag 676
```

Search completed: July 20, 2002, 02:41:40
Job time: 7818 sec

OM of: US-09-836-410A-1 to: GenEmbl:* out_format : pfs

Date: Jul 20, 2002 4:49 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_115032_29702/app_query.fasta_1.660
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n15to50.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=15 -MAXLEN=50
-USER=US09836410_@CGN1_1_4586 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2148.830000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_pat:AX081685	+	44.00	83.00	2.9e+04	49	! AX081685 Sequence 34 from Patent
gb_pat:AX135674	+	44.00	83.00	2.9e+04	49	! AX135674 Sequence 15 from Patent
gb_pat:I27765	+	43.00	81.77	3.4e+04	48	! I27765 Sequence 7 from patent US
gb_pat:AX159092	-	43.00	81.41	3.6e+04	50	! AX159092 Sequence 2420 from Pate
gb_pat:AR126194	+	42.00	80.93	3.8e+04	45	! AR126194 Sequence 212 from paten
gb_pat:A70028	+	40.50	79.42	4.6e+04	42	! A70028 Sequence 11 from Patent E
gb_pat:AX221610	+	40.50	78.24	5.3e+04	48	! AX221610 Sequence 7052 from Pate
gb_pr:HUMRPY59	-	40.00	77.54	5.8e+04	48	! D28400 Homo sapiens mRNA for rib
gb_pat:AX147194	-	40.00	77.18	6.1e+04	50	! AX147194 Sequence 32 from Patent
gb_pat:AX222245	+	39.50	76.83	6.4e+04	48	! AX222245 Sequence 7687 from Pate
gb_pat:AX111996	+	39.00	80.28	4.1e+04	30	! AX111996 Sequence 9 from Patent
gb_pat:AR123920	-	39.00	76.32	6.8e+04	47	! AR123920 Sequence 4 from patent
gb_pat:I29948	-	39.00	75.77	7.3e+04	50	! I29948 Sequence 12 from patent U
gb_pat:AX223522	+	38.50	75.42	7.6e+04	48	! AX223522 Sequence 8964 from Pate
gb_pat:AX229410	+	38.50	75.42	7.6e+04	48	! AX229410 Sequence 2782 from Pate
gb_pat:AX043391	-	38.00	80.48	4.0e+04	25	! AX043391 Sequence 957 from Paten
gb_pl:S71800S1	-	38.00	80.13	4.2e+04	26	! S71800 {autonomous Activator ele
gb_pat:AX063334	+	38.00	76.55	6.6e+04	39	! AX063334 Sequence 43 from Patent
gb_pat:A11683	+	38.00	75.69	7.4e+04	43	! A11683 Delta-EA1A nucleotide fra
gb_pat:A11684	-	38.00	75.69	7.4e+04	43	! A11684 Delta-EA1A nucleotide fra
gb_pat:A12236	+	38.00	75.69	7.4e+04	43	! A12236 Nucleotide sequence 20 fr
gb_pat:A25147	+	38.00	75.69	7.4e+04	43	! A25147 Artificial DNA for aproti
gb_pat:A25148	+	38.00	75.69	7.4e+04	43	! A25148 Artificial DNA for aproti
gb_pat:BD006271	+	38.00	75.10	8.0e+04	46	! BD006271 Artificial antibody pol
gb_pat:I29950	-	38.00	75.10	8.0e+04	46	! I29950 Sequence 14 from patent U
gb_pat:AR053336	+	38.00	74.91	8.2e+04	47	! AR053336 Sequence 3 from patent
gb_pat:AR053340	+	38.00	74.91	8.2e+04	47	! AR053340 Sequence 7 from patent
gb_pat:AR077201	+	38.00	74.91	8.2e+04	47	! AR077201 Sequence 17 from patent
gb_pat:AX003441	+	38.00	74.91	8.2e+04	47	! AX003441 Sequence 21 from Patent
gb_pat:AX112276	+	38.00	74.91	8.2e+04	47	! AX112276 Sequence 5 from Patent
gb_pat:E38122	+	38.00	74.91	8.2e+04	47	! E38122 DNA elongation and analys
gb_pr:HSMTCP1A	-	38.00	74.91	8.2e+04	47	! Z24455 H.sapiens MTCP-1 gene. 9/
gb_pat:AX229252	+	38.00	74.72	8.4e+04	48	! AX229252 Sequence 2624 from Pate
gb_pat:A12237	-	38.00	74.54	8.6e+04	49	! A12237 Nucleotide sequence 21 fr
gb_pat:AX157759	-	38.00	74.36	8.7e+04	50	! AX157759 Sequence 1087 from Pate
gb_pat:AX157761	-	38.00	74.36	8.7e+04	50	! AX157761 Sequence 1089 from Pate
gb_pat:AX158154	-	38.00	74.36	8.7e+04	50	! AX158154 Sequence 1482 from Pate
gb_pat:AR161538	+	37.50	75.19	7.9e+04	42	! AR161538 Sequence 38 from patent
gb_pat:AX229539	+	37.50	74.02	9.1e+04	48	! AX229539 Sequence 2911 from Pate
gb_pat:AX248586	-	37.00	77.17	6.1e+04	31	! AX248586 Sequence 665 from Paten
gb_pat:AX249525	-	37.00	77.17	6.1e+04	31	! AX249525 Sequence 1604 from P
gb_pat:AX183714	-	37.00	76.62	6.6e+04	33	! AX183714 Sequence 1467 from P
gb_pat:AX137735	+	37.00	75.85	7.2e+04	36	! AX137735 Sequence 8 from Pate
gb_pat:AX207592	+	37.00	75.85	7.2e+04	36	! AX207592 Sequence 1 from Pate
gb_pat:AR004318	+	37.00	75.61	7.5e+04	37	! AR004318 Sequence 1 from pate

seq_name: gb_pat:AX081685

seq_documentation_block:
LOCUS AX081685 49 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 34 from Patent WO0109347.
ACCESSION AX081685
VERSION AX081685.1 GI:13170509
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 49)
AUTHORS Hogrefe,H.H., Cline,J.M., Hansen,C.J. and Borns,M.C.
TITLE Archaeal replication accessory factors and methods of use
JOURNAL Patent: WO 0109347-A 34 08-FEB-2001;
STRATAGENE (US)
FEATURES Location/Qualifiers
source 1. .49
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT 18 a 8 c 14 g 9 t
ORIGIN

alignment_scores:
Quality: 44.00 Length: 16
Ratio: 3.385 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:
US-09-836-410A-1 x AX081685 ..

Align seg 1/1 to: AX081685 from: 1 to: 49

```
    359 AspAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLys 374
        ||||||:::::::::|||   |||      |||:::|||||||:::|||
      1 GACGACGACAAGATGGGTGTCCCAATTGGTGAGATTATACCAAGAAAA 48
```

seq_name: gb_pat:AX135674

seq_documentation_block:
LOCUS AX135674 49 bp DNA linear PAT 29-MAY-2001
DEFINITION Sequence 15 from Patent WO0132922.
ACCESSION AX135674
VERSION AX135674.1 GI:14271944
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 49)
AUTHORS Sorge,J.A.
TITLE Methods for detection of a target nucleic acid sequence
JOURNAL Patent: WO 0132922-A 15 10-MAY-2001;
STRATAGENE (US)
FEATURES Location/Qualifiers
source 1. .49
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic pCAL-n-EK cloning synthetic
oligonucleotide"
BASE COUNT 18 a 8 c 14 g 9 t
ORIGIN

alignment_scores:
Quality: 44.00 Length: 16

```
            Ratio:  3.385                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AX135674   ..

 Align seg 1/1  to: AX135674  from: 1  to: 49

     359 AspAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLys 374
         ||||||::::::::::|||   |||      |||:::|||||||:::|||
       1 GACGACGACAAGATGGGTGTCCCAATTGGTGAGATTATACCAAGAAAA 48

seq_name: gb_pat:I27765

seq_documentation_block:
LOCUS       I27765                  48 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 7 from patent US 5567602.
ACCESSION   I27765
VERSION     I27765.1  GI:1818541
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Clark,J.M., Shoemaker,K.R. and Warne,R.L.
  TITLE     Recombinant production of chymase
  JOURNAL   Patent: US 5567602-A 7 22-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="unknown"
BASE COUNT       16 a     11 c     14 g      7 t
ORIGIN


alignment_scores:
            Quality:   43.00             Length:     10
              Ratio:  4.778               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-836-410A-1 x I27765   ..

 Align seg 1/1  to: I27765  from: 1  to: 48

     356 LysAspAspAspAspGluGluIleGlyGly 365
         ||||||||||||||:::   |||||||||
      13 AAGGACGACGATGACAAGATCATCGGGGGC 42

seq_name: gb_pat:AX159092

seq_documentation_block:
LOCUS       AX159092                50 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 2420 from Patent WO0140521.
ACCESSION   AX159092
VERSION     AX159092.1  GI:14540423
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: WO 0140521-A 2420 07-JUN-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    25. .26
                     /note="Nucleotide deleted between bases 25 and 26
                     Accession number cg39584802"
     misc_feature    26
                     /note="2 of 2 allelic variants (2419 is other entry)"
BASE COUNT       10 a     17 c     14 g      9 t
ORIGIN


alignment_scores:
            Quality:   43.00             Length:      9
              Ratio:  4.778               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-836-410A-1 x AX159092/rev   ..

 Align seg 1/1  to reverse of: AX159092  from: 1  to: 50

      17 ArgGlyLeuValProArgLysLeuPro 25
         |||||||||||||||||||:::|||||||
      43 AGAGGGCTGGTCCCACGTCAGCTGCCG 17

seq_name: gb_pat:AR126194

seq_documentation_block:
LOCUS       AR126194                45 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 212 from patent US 6180084.
ACCESSION   AR126194
VERSION     AR126194.1  GI:14112787
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Ruoslahti,E. and Pasqualini,R.
  TITLE     NGR receptor and methods of identifying tumor homing molecules that
            home to angiogenic vasculature using same
  JOURNAL   Patent: US 6180084-A 212 30-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="unknown"
BASE COUNT       10 a     13 c     15 g      7 t
ORIGIN


alignment_scores:
            Quality:   42.00             Length:     14
              Ratio:  3.818               Gaps:      0
 Percent Similarity:  78.571   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AR126194   ..

 Align seg 1/1  to: AR126194  from: 1  to: 45

      91 AspAspGlyLysGluGluProProThrThrLeuLeuTrpVal 104
         :::|||   :::|||   ||||||:::|||:::   ||||||
       2 GAAGATCTAGAGGAACCACCGCCTGTGACGGTCAGCTGGGTC 43

seq_name: gb_pat:A70028

seq_documentation_block:
LOCUS       A70028                  42 bp    DNA     linear   PAT 07-MAY-1999
DEFINITION  Sequence 11 from Patent EP0832972.
ACCESSION   A70028
VERSION     A70028.1  GI:4774476
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Silke,N., Lerch,K. and Muheim,A.
  TITLE     Cloning, expression and production of tasty peptides
  JOURNAL   Patent: EP 0832972-A 11 01-APR-1998;
            GIVAUDAN ROURE INT (CH)
```

```
FEATURES             Location/Qualifiers
     source    .     1. .42
                    /organism="unidentified"
                    /db_xref="taxon:32644"
BASE COUNT        8 a     14 c      5 g     15 t
ORIGIN


alignment_scores:
            Quality:   40.50             Length:     15
              Ratio:  3.375               Gaps:      1
 Percent Similarity:  80.000   Percent Identity: 53.333

alignment_block:
 US-09-836-410A-1 x A70028   ..

 Align seg 1/1  to: A70028  from: 1  to: 42

     243 PheIleGluIleThrAspAspGlnPheAspPheHisThrTyrCys 257
         |||:::|||   |||||:::   |||||||||   :::::|||
       1 TTCTTAGAA...ACCGATTCTCTCTTCGACTTCGATTCCCACTGC 42

seq_name: gb_pat:AX221610

seq_documentation_block:
LOCUS       AX221610                 48 bp    mRNA    linear   PAT 07-SEP-2001
DEFINITION  Sequence 7052 from Patent WO0159103.
ACCESSION   AX221610
VERSION     AX221610.1  GI:15549334
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Blatt,L., Mcswiggen,J. and Chowrira,B.M.
  TITLE     Method and reagent for the modulation and diagnosis of cd20 and
            nogo gene expression
  JOURNAL   Patent: WO 0159103-A 7052 16-AUG-2001;
            RIBOZYME PHARMACEUTICALS, INC. (US) ; Blatt, Lawrence (US) ;
            McSwiggen, James (US) ; Chowrira, Bharat M. (US)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Nucleic Acid"
BASE COUNT       15 a     11 c     12 g     10 t
ORIGIN


alignment_scores:
            Quality:   40.50             Length:     19
              Ratio:  2.893               Gaps:      1
 Percent Similarity:  73.684   Percent Identity: 47.368

alignment_block:
 US-09-836-410A-1 x AX221610   ..

 Align seg 1/1  to: AX221610  from: 1  to: 48

     381 ProLeuGluGluAlaIleLysPheLeuThrProLeuLysAsnLeuValLy 397
         ||||||||||||              |||||:::|||:::::|||::
       4 CCGTTGGAGGAA.............ACTCCCTTCAAGGACATCGTCCG 38

     397 sAsnLys 399
         ::::|||
      39 GGATAAA 45

seq_name: gb_pr:HUMRPY59

seq_documentation_block:
LOCUS       HUMRPY59                 48 bp    mRNA    linear   PRI 27-SEP-2001
DEFINITION  Homo sapiens mRNA for ribosomal protein L24, partial cds.
ACCESSION   D28400
VERSION     D28400.1  GI:461272
KEYWORDS    .
SOURCE      Homo sapiens lymphoma cell_line:U937 cDNA to mRNA,
            clone_lib:U937/pKA1 clone:HP00302.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Kato,S., Sekine,S., Oh,S.W., Kim,N.S., Umezawa,Y., Abe,N.,
            Yokoyama-Kobayashi,M. and Aoki,T.
  TITLE     Construction of a human full-length cDNA bank
  JOURNAL   Gene. 150 (2), 243-250 (1994)
  MEDLINE   95121910
REFERENCE   2  (bases 1 to 48)
  AUTHORS   Kato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1994) Seishi Kato, Research Institute of National
            Rehabilitation Center for the Disabled, Department of
            Rehabilitation Engineering; 4-1 Namiki, Tokorozawa, Saitama
            359-8555, Japan (E-mail:seishi@rehab.go.jp,
            Tel:81-42-995-3100(ex.2568), Fax:81-42-995-3132)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HP00302"
                     /cell_line="U937"
                     /tissue_type="lymphoma"
                     /clone_lib="U937/pKA1"
     5'UTR           1. .42
     variation       1. .10
                     /replace="ttttctttttc"
     variation       1. .10
                     /replace="ctttttttttc"
     variation       1. .10
                     /replace="ctttttttttttc"
     variation       1. .10
                     /replace="ttttttctttttc"
     variation       1. .10
                     /replace="ctttctttttc"
     CDS             43. .>48
                     /codon_start=1
                     /product="ribosomal protein L24"
                     /protein_id="BAA05766.1"
                     /db_xref="GI:4433237"
                     /translation="MK"
BASE COUNT        5 a     13 c     10 g     20 t
ORIGIN


alignment_scores:
            Quality:   40.00             Length:     15
              Ratio:  2.857               Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 46.667

alignment_block:
 US-09-836-410A-1 x HUMRPY59/rev   ..

 Align seg 1/1  to reverse of: HUMRPY59  from: 1  to: 48

     578 ValAsnGlyAspSerSerAlaGluThrGluGluLeuAlaAsnGlu 592
         ::::::|||||||||||||:::|||   ::::::::::|||:::|||
      48 CTTCATGGCGACAGCTCCACGGAAAGACAAAAGATGGCGAAAGAA 4

seq_name: gb_pat:AX147194

seq_documentation_block:
LOCUS       AX147194                 50 bp    DNA     linear   PAT 08-JUN-2001
DEFINITION  Sequence 32 from Patent WO0136682.
ACCESSION   AX147194
VERSION     AX147194.1  GI:14346365
KEYWORDS    .
SOURCE      synthetic construct.
```

```
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Chenchik,A., Munishkin,A. and Simonenko,P.
  TITLE     Long oligonucleotide arrays
  JOURNAL   Patent: WO 0136682-A 32 25-MAY-2001;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="synthetic oligonucleotide"
BASE COUNT       17 a      6 c     16 g     11 t
ORIGIN


alignment_scores:
            Quality:   40.00             Length:      9
              Ratio:  4.444               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
 US-09-836-410A-1 x AX147194/rev   ..

 Align seg 1/1  to reverse of: AX147194  from: 1  to: 50

     428 AlaIleAspSerSerHisProTrpLeu 436
         |||:::::::::|||||||||||||||
      44 GCACTTTCTCAAAGTCACCCTTGGCTG 18

seq_name: gb_pat:AX222245

seq_documentation_block:
LOCUS       AX222245                 48 bp    mRNA    linear   PAT 07-SEP-2001
DEFINITION  Sequence 7687 from Patent WO0159103.
ACCESSION   AX222245
VERSION     AX222245.1  GI:15549969
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Blatt,L., Mcswiggen,J. and Chowrira,B.M.
  TITLE     Method and reagent for the modulation and diagnosis of cd20 and
            nogo gene expression
  JOURNAL   Patent: WO 0159103-A 7687 16-AUG-2001;
            RIBOZYME PHARMACEUTICALS, INC. (US) ; Blatt, Lawrence (US) ;
            McSwiggen, James (US) ; Chowrira, Bharat M. (US)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Nucleic Acid"
BASE COUNT       12 a     16 c     11 g      9 t
ORIGIN


alignment_scores:
            Quality:   39.50             Length:     18
              Ratio:  3.038               Gaps:      1
 Percent Similarity:  72.222   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AX222245   ..

 Align seg 1/1  to: AX222245  from: 1  to: 48

     380 ThrProLeuGluGluAlaIleLysPheLeuThrProLeuLysAsnLeuVa 396
         ||||||||||||||||              ||||||:::|||:::::::||
       1 ACACCACTGGAGGAA.............ACTCCCTTCAAGGACATCGT 35

     396 lLys 397
         |:::
```

```
      36 CCGG 39

seq_name: gb_pat:AX111996

seq_documentation_block:
LOCUS       AX111996                 30 bp    DNA     linear   PAT 01-MAY-2001
DEFINITION  Sequence 9 from Patent WO0125439.
ACCESSION   AX111996
VERSION     AX111996.1  GI:13938904
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Bonello,J.F., Rogowsky,P. and Perez,P.
  TITLE     Plant seed endosperm-specific promoter
  JOURNAL   Patent: WO 0125439-A 9 12-APR-2001;
            Biogemma (FR)
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="SEQUENCE DESCRIPTION artificielle:oligonucleotide"
BASE COUNT        9 a      6 c      6 g      9 t
ORIGIN


alignment_scores:
            Quality:   39.00             Length:     10
              Ratio:  4.333               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-836-410A-1 x AX111996   ..

 Align seg 1/1  to: AX111996  from: 1  to: 30

     230 GlyGluAlaLeuLysLysCysHisGluIle 239
         ||||||||||||   :::|||||||:::|||
       1 GGGGAAGCTTTACATTCTTGCCATAACATA 30

seq_name: gb_pat:AR123920

seq_documentation_block:
LOCUS       AR123920                 47 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 4 from patent US 6171823.
ACCESSION   AR123920
VERSION     AR123920.1  GI:14109281
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Woldike,H.Fabricius. and Hastrup,S.
  TITLE     Process for producing extracellular proteins in bacteria
  JOURNAL   Patent: US 6171823-A 4 09-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="unknown"
BASE COUNT        6 a     18 c     12 g     11 t
ORIGIN


alignment_scores:
            Quality:   39.00             Length:      9
              Ratio:  4.333               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
 US-09-836-410A-1 x AR123920/rev   ..

 Align seg 1/1  to reverse of: AR123920  from: 1  to: 47
```

```
     355 LysLysAspAspAspAspGluGluIle 363
         :::|||||||||||||||||:::|||:::
      27 GAGAAGGACGACGATGATAAAGAGGTC 1

seq_name: gb_pat:I29948

seq_documentation_block:
LOCUS       I29948                    50 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 12 from patent US 5578478.
ACCESSION   I29948
VERSION     I29948.1  GI:1820739
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Rambosek,J., Piddington,C.S., Kovacevich,B.R., Young,K.D. and
            Denome,S.A.
  TITLE     Recombinant DNA encoding a desulfurization biocatalyst
  JOURNAL   Patent: US 5578478-A 12 26-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="unknown"
BASE COUNT       17 a     13 c     12 g      8 t
ORIGIN


alignment_scores:
            Quality:   39.00             Length:      9
              Ratio:  4.875               Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 66.667

alignment_block:
 US-09-836-410A-1 x I29948/rev   ..

 Align seg 1/1  to reverse of: I29948  from: 1  to: 50

     553 AlaSerCysHisLysLeuPheProTyr 561
         |||:::||||||   |||||||||:::
      43 GCGGCTTGTCATGGTCTGTTTCCTCAT 17

seq_name: gb_pat:AX223522

seq_documentation_block:
LOCUS       AX223522                  48 bp    mRNA    linear   PAT 07-SEP-2001
DEFINITION  Sequence 8964 from Patent WO0159103.
ACCESSION   AX223522
VERSION     AX223522.1  GI:15551246
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Blatt,L., Mcswiggen,J. and Chowrira,B.M.
  TITLE     Method and reagent for the modulation and diagnosis of cd20 and
            nogo gene expression
  JOURNAL   Patent: WO 0159103-A 8964 16-AUG-2001;
            RIBOZYME PHARMACEUTICALS, INC. (US) ; Blatt, Lawrence (US) ;
            McSwiggen, James (US) ; Chowrira, Bharat M. (US)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Nucleic Acid"
BASE COUNT       13 a     13 c     12 g     10 t
ORIGIN


alignment_scores:
            Quality:   38.50             Length:     21
              Ratio:  2.406               Gaps:       1
 Percent Similarity:  76.190   Percent Identity: 38.095

alignment_block:
 US-09-836-410A-1 x AX223522   ..

 Align seg 1/1  to: AX223522  from: 1  to: 48

     380 ThrProLeuGluGluAlaIleLysPheLeuThrProLeuLysAsnLeuVa 396
         :::|||||||||||||              ||||||:::|||:::::||
       1 AGCCCATTGGAGGAA..............ACTCCCTTCAAGGACATCGT 35

     396 lLysAsnLysIle 400
         |::::::::::::
      36 CCGGGATAATCTG 48

seq_name: gb_pat:AX229410

seq_documentation_block:
LOCUS       AX229410                  48 bp    mRNA    linear   PAT 10-SEP-2001
DEFINITION  Sequence 2782 from Patent WO0157206.
ACCESSION   AX229410
VERSION     AX229410.1  GI:15558551
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Fattaey,A.R., Jarvis,T., Mcswiggen,J., Booher,R.N. and Holman,P.S.
  TITLE     Method and reagent for the inhibition of checkpoint kinase-1 (chk
            1) enzyme
  JOURNAL   Patent: WO 0157206-A 2782 09-AUG-2001;
            RIBOZYME PHARMACEUTICALS, INC. (US) ; Fattaey, Ali R. (US)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT       11 a     14 c     14 g      9 t
ORIGIN


alignment_scores:
            Quality:   38.50             Length:     21
              Ratio:  2.567               Gaps:       1
 Percent Similarity:  71.429   Percent Identity: 38.095

alignment_block:
 US-09-836-410A-1 x AX229410   ..

 Align seg 1/1  to: AX229410  from: 1  to: 48

     380 ThrProLeuGluGluAlaIleLysPheLeuThrProLeuLysAsnLeuVa 396
         ||||||   ||||||             ||||||:::|||:::::||
       1 ACCCCTGCGGAGGAA..............ACTCCCTTCAAGGACATCGT 35

     396 lLysAsnLysIle 400
         |::::::::::::
      36 CCGGGATGAGTTG 48
```

OM of: US-09-836-410A-1 to: N_Geneseq_032802:* out_format : pfs

Date: Jul 20, 2002 4:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_115033_29721/app_query.fasta_1.660
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n15to50.rng
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=15 -MAXLEN=50
-USER=US09836410_@CGN1_1_188 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 200.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH27115	+	44.00	88.85	1.3e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF57073	+	44.00	88.85	1.3e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT47976	+	43.00	87.32	1.6e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA08530	-	43.00	86.88	1.7e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI75479	-	43.00	86.88	1.7e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL34450	+	42.50	86.00	1.9e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX88379	+	42.00	86.25	1.9e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX23741	+	42.00	86.25	1.9e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ67157	+	42.00	85.78	2.0e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH75873	-	41.00	85.48	2.1e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX05908	+	41.00	84.02	2.5e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX15480	+	41.00	84.02	2.5e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ67285	-	41.00	84.02	2.5e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:AAN92016	-	41.00	83.37	2.7e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL34455	+	41.00	83.37	2.7e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABK07052	+	40.50	82.92	2.8e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA95836	-	40.00	89.43	1.2e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:AAQ05380	-	40.00	84.82	2.2e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ55044	+	40.00	84.26	2.4e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH75874	-	40.00	83.72	2.6e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH77532	-	40.00	81.61	3.4e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABK07687	+	39.50	81.16	3.6e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF03584	+	39.00	85.65	2.0e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF04032	+	39.00	85.65	2.0e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC63569	-	39.00	83.65	2.6e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT13788	-	39.00	80.51	3.9e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ96614	+	39.00	80.28	4.0e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ55127	-	39.00	79.85	4.2e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:AAT58976	-	39.00	79.85	4.2e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX15621	-	39.00	79.85	4.2e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL27979	-	39.00	79.85	4.2e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH97568	+	38.50	79.41	4.4e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABK08964	+	38.50	79.41	4.4e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:AAT80426	+	38.50	79.19	4.6e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC96528	-	38.00	85.48	2.1e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV18765	+	38.00	81.59	3.4e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF24615	+	38.00	80.74	3.7e+0		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV51145	-	38.00	80.47	3.9e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ87311	-	38.00	80.21	4.0e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA08786	-	38.00	79.45	4.4e+04		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC73547	+	38.00	79.21	4.6e		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ55129	-	38.00	78.98	4.7e		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:AAT58978	-	38.00	78.98	4.7e		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX15623	-	38.00	78.98	4.7e		
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX05505	+	38.00	78.98	4.7e		

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH27115

seq_documentation_block:
ID AAH27115 standard; DNA; 49 BP.
XX
AC AAH27115;
XX
DT 06-AUG-2001 (first entry)
XX
DE PCR primer used for FEN endonuclease gene cloning.
XX
KW Cleavage structure; target sequence detection; flap endonuclease;
KW FEN; PCR primer; ss.
XX
OS Pyrococcus furiosus.
XX
PN WO200132922-A2.
XX
PD 10-MAY-2001.
XX
PF 27-OCT-2000; 2000WO-US29663.
XX
PR 29-OCT-1999; 99US-0430692.
XX
PA (STRA-) STRATAGENE.
XX
PI Sorge JA;
XX
DR WPI; 2001-328805/34.
XX
PT The labelling of nucleic acids for their detection and quantification
PT comprises the formation of a cleavage structure and its cleavage with a
PT five' exonuclease-1 or flap endonuclease-1 -
XX
PS Example 2; Page 55; 81pp; English.
XX
CC This invention relates to a method for generating a signal indicative of
CC the presence of a target nucleic acid sequence in a sample. The method
CC comprises the formation of a cleavage structure through the incubation of
CC a sample comprising a target nucleic acid sequence and a nucleic acid
CC polymerase and cleaving the cleavage structure with a 5' exonuclease-1 or
CC flap endonuclease (FEN) to generate the signal. The method is used for
CC the detection and quantification of a target nucleic acid sequence. The
CC present sequence represents a PCR primer used to amplify the Pyrococcus
CC furiosus FEN endonuclease gene sequence. The PCR product is used in an
CC example illustrating the method of the invention.
XX
SQ Sequence 49 BP; 18 A; 8 C; 14 G; 9 T; 0 other;

alignment_scores:
Quality: 44.00 Length: 16
Ratio: 3.385 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:
US-09-836-410A-1 x AAH27115 ..

Align seg 1/1 to: AAH27115 from: 1 to: 49

```
    359 AspAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLys 374
        ||||||::::::::|||   |||      |||:::||||||::::|||
      1 GACGACGACAAGATGGGTGTCCCAATTGGTGAGATTATACCAAGAAAA 48
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF57073

seq_documentation_block:

```
ID   AAF57073 standard; DNA; 49 BP.
XX
AC   AAF57073;
XX
DT   14-MAY-2001  (first entry)
XX
DE   P. furiosus FEN-1 gene amplifying forward primer.
XX
KW   MCM; minichromosome maintenance protein; archaeal polypeptide; PCNA;
KW   RFC-P38; RFC-P55; RFA; CDC6; FEN-1; dUTPase; ligase; helicase dna2;
KW   nucleic acid amplification; polymerase chain reaction; PCR primer; ss.
XX
OS   Pyrococcus furiosus.
XX
PN   WO200109347-A2.
XX
PD   08-FEB-2001.
XX
PF   28-JUL-2000; 2000WO-US20532.
XX
PR   30-JUL-1999;   99US-0146580.
XX
PA   (STRA-) STRATAGENE.
XX
PI   Hogrefe HH,  Cline JM,  Hansen CJ,  Borns MC;
XX
DR   WPI; 2001-182959/18.
XX
PT   Composition for improving nucleic acid polymerase reactions, useful
PT   e.g. in synthesis or amplification, contains at least one archaeal
PT   accessory protein -
XX
PS   Disclosure; Page 29; 147pp; English.
XX
CC   The invention provides a composition (A) for enhancing nucleic acid
CC   polymerase reactions that comprises an archaeal MCM (minichromosome
CC   maintenance protein) and at least one of the archaeal polypeptides (PCNA,
CC   RFC-P38 or -P55, RFA, CDC6, FEN-1, dUTPase, ligase, helicase dna2, or
CC   helicases 2-8). (A) And similar compositions containing different
CC   combinations of accessory proteins, are used to improve performance of
CC   synthesis, amplification, mutagenizing, labeling and detecting reactions,
CC   e.g. for gene characterization, cloning, detection of allelic variants,
CC   diagnosis and screening for disease, particularly where done by
CC   polymerase chain reaction (PCR). Some of the proteins also stabilize
CC   duplexes during polymerase reactions or improve exonuclease reactions,
CC   for example RFA also improves specificity of nucleic acid/protein
CC   interaction and PCNA improves polymerase-mediated repair processes and
CC   hybridization reactions. Nucleic acids encoding the archaeal polypeptides
CC   are used for recombinant production of proteins, and fragments of the
CC   nucleic acid as probes and primers for screening related sequences. The
CC   accessory proteins increase accuracy and efficiency of polymerase
CC   reactions, allow use of lower denaturation and extension temperatures
CC   (possibly isothermal processing), and improve synthesis of long targets.
CC   Sequences AAF57041-74 represent PCR primers for amplifying various
CC   P. furiosus genes, for the production of accessory factors.
XX
SQ   Sequence 49 BP; 18 A; 8 C; 14 G; 9 T; 0 other;


alignment_scores:
            Quality:   44.00              Length:     16
              Ratio:  3.385                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AAF57073   ..

 Align seg 1/1  to: AAF57073  from: 1  to: 49

     359 AspAspGluGluIleGlyGlyProLysGluGluLeuIleProGluLys 374
         ||||||::::::::::|||   |||      |||:::|||||||:::|||
       1 GACGACGACAAGATGGGTGTCCCAATTGGTGAGATTATACCAAGAAAA 48

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT47976

seq_documentation_block:
ID   AAT47976 standard; DNA; 48 BP.
XX
AC   AAT47976;
XX
DT   14-MAR-1997  (first entry)
XX
DE   5' primer 1 for amplifying human chymase gene.
XX
KW   Wild type; human; chymase; fusion protein; proteolytically; unstable;
KW   inactive; cleavage; transition mutation; antibody; inhibitor; primer;
KW   PCR; polymerase chain reaction; amplification; ss.
XX
OS   Synthetic.
XX
PN   US5567602-A.
XX
PD   22-OCT-1996.
XX
PF   12-AUG-1992;   92US-0929198.
XX
PR   12-AUG-1992;   92US-0929198.
XX
PA   (ARRI-) ARRIS PHARM CORP.
XX
PI   Clark JM,  Shoemaker KR,  Warne RL;
XX
DR   WPI; 1996-484995/48.
XX
PT   New recombinant production of active chymase - using a fusion
PT   protein of a non-chymase protein attached to chymase which is
PT   refolded; for producing antibodies and screening for inhibitors
XX
PS   Examples; Column 21-22; 31pp; English.
XX
CC   The primers AAT47976-9 were used to amplify the sequence encoding the
CC   human chymase catalytic domain.  A first amplification was carried out
CC   using primers AAT47976 and AAT47979.  A second amplification was then
CC   carried out with primers AAT47977 and AAT47979. Positive transformants
CC   were screened by PCR using primers AAT47978-9.  The amplifications
CC   resulted in a chymase protein which contained the myelin-associated
CC   glycoprotein (MAG) and the 8 amino acid FLAG epitope/cleavage site fused
CC   to the N-terminus of the chymase catalytic domain.  This fusion allows
CC   easier purification of the chymase from a mammalian expression system.
CC   The chymase portion of the fusion was derived from the wild type gene
CC   (AAT47973) or the clones 3 (AAT47974) and 7 (AAT47975).  The use of a
CC   chymase fusion is because chymase produced by itself is proteolytically
CC   active so tends to be unstable in the cell.  The fusion portion of the
CC   protein renders the chymase inactive until it removed by cleavage.  The
CC   purified chymase can be used to generate chymase-specific antibodies and
CC   to identify chymase inhibitors.
XX
SQ   Sequence 48 BP; 16 A; 11 C; 14 G; 7 T; 0 other;


alignment_scores:
            Quality:   43.00              Length:     10
              Ratio:  4.778                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-836-410A-1 x AAT47976   ..

 Align seg 1/1  to: AAT47976  from: 1  to: 48

     356 LysAspAspAspAspGluGluIleGlyGly 365
         ||||||||||||||::   |||||||||
      13 AAGGACGACGATGACAAGATCATCGGGGGC 42

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA08530
```

```
seq_documentation_block:
ID   AAA08530 standard; DNA; 50 BP.
XX
AC   AAA08530;
XX
DT   19-JUL-2000  (first entry)
XX
DE   Oligonucleotide encoding C-terminal portion of the alpha factor signal
DE   sequence and mutated trypsinogen leader sequence.
XX
KW   C-terminal alpha factor signal sequence; trypsinogen; leader sequence;
KW   analogue; mutated bovine trypsinogen; recombinant protein production;
KW   inactive zymogen; ss.
XX
OS   Synthetic.
XX
PN   WO200017332-A1.
XX
PD   30-MAR-2000.
XX
PF   15-SEP-1999;   99WO-US21047.
XX
PR   21-SEP-1998;   98US-0101213.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Hanquier JM,  Hershberger CL,  Desplancq D,  Larson JL,  Rosteck PR;
XX
DR   WPI; 2000-283565/24.
XX
PT   New trypsinogen analog useful for the production of recombinant trypsin
PT   has a modified leader sequence not cleavable by trypsin or trypsin-like
PT   enzymes
XX
PS   Example 1; Page 29; 56pp; English.
XX
CC   Trypsinogen was fused directly to the C-terminus of the alpha factor
CC   without a Glu-Ala-Glu-Ala linker peptide. Oligonucleotides AAA08529-30
CC   encoding a C-terminal portion of the alpha factor signal sequence and
CC   the Val(Asp)5 leader sequence were synthesized. The wild type bovine
CC   trypsinogen was mutated to destroy the trypsin cleavage site. The lys
CC   residue present in the leader sequence of the native bovine trypsinogen
CC   protein was mutated to an Asp residue. The specification claims an
CC   isolated trypsinogen analogue comprising a protein having trypsin
CC   activity and a leader sequence having at least two amino acids which
CC   are not Lys or Arg. A recombinantly produced trypsin (AAY91926) is also
CC   claimed. The trypsin derived from the recombinant trypsinogen is useful
CC   for the characterization of other proteins, and in the manufacture of
CC   other recombinant bioproducts, for example to cleave leader sequences
CC   from small recombinant proteins expressed initially as fusion proteins.
CC   The present method provides for expression of an inactive zymogen form
CC   that is soluble and properly folded yet is not activated until after
CC   purification from fermentation broth or cell extracts.  This is
CC   accomplished through the expression of a single chain trypsinogen
CC   analogue where the leader sequence is modified such that it lacks a
CC   trypsin-like enzyme cleavage site.  Specifically the trypsinogen
CC   analogues of the present invention lack a lysine or arginine in the
CC   N-terminal leader sequence of the protein to prevent auto-activation or
CC   activation by endougenous host cell enzymes.
XX
SQ   Sequence 50 BP; 12 A; 17 C; 6 G; 15 T; 0 other;


alignment_scores:
           Quality:   43.00              Length:      12
             Ratio:  4.300                Gaps:       0
 Percent Similarity:  83.333   Percent Identity: 58.333

alignment_block:
 US-09-836-410A-1 x AAA08530/rev   ..

 Align seg 1/1  to reverse of: AAA08530  from: 1  to: 50

      354 LysLysLysAspAspAspAspGluGluIleGlyGly 365
          |||:::   |||||||||||||:::   :::||||||
       49 AAAAGAGTCGACGATGATGACGATATCGTTGGAGGT 14

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI75479

seq_documentation_block:
ID   AAI75479 standard; DNA; 50 BP.
XX
AC   AAI75479;
XX
DT   09-NOV-2001  (first entry)
XX
DE   Human silent SNP containing nucleic acid SEQ:2420.
XX
KW   Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW   protein therapy; vaccine; probe; diagnostic assay; detection;
KW   quantitation; restorative therapy; polymorphic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200140521-A2.
XX
PD   07-JUN-2001.
XX
PF   30-NOV-2000; 2000WO-US32758.
XX
PR   30-NOV-1999;   99US-0168138.
PR   29-NOV-2000; 2000US-0726173.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2001-356160/37.
XX
PT   Polymorphic nucleic acid sequences, useful in genetic testing and
PT   therapy -
XX
PS   Claim 1; Page 792; 2653pp; English.
XX
CC   AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC   sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC   AAM53114 to AAM53329 represent peptides related to human polymorphic
CC   polynucleotide sequences. The sequences can be used in gene and protein
CC   therapy, and in vaccine production. (I) and the polypeptides encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression of polymorphic polypeptides.
CC   For example, (I) may be used to treat disorders by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patients own production of polypeptide. Additionally, (I) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids
CC   in samples, and therefore which patients may be in need of restorative
CC   therapy. The polypeptides encoded by (I) may be used as antigens in the
CC   production of antibodies specific for polymorphic polypeptides. The
CC   antibodies may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of polymorphic polypeptides in samples.
XX
SQ   Sequence 50 BP; 10 A; 17 C; 14 G; 9 T; 0 other;


alignment_scores:
           Quality:   43.00              Length:       9
             Ratio:  4.778                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-836-410A-1 x AAI75479/rev   ..

 Align seg 1/1  to reverse of: AAI75479  from: 1  to: 50
```

```
     17 ArgGlyLeuValProArgLysLeuPro 25
        ||||||||||||||||||:::||||||
     43 AGAGGGCTGGTCCCACGTCAGCTGCCG 17

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL34450

seq_documentation_block:
ID   AAL34450 standard; DNA; 50 BP.
XX
AC   AAL34450;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Human SNP oligonucleotide #7658.
XX
KW   Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
KW   neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
KW   amyloid protein; angiopoietin; apoptosis related protein; cadherin;
KW   cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
KW   complement related protein; cytochrome; kinesin; cytokine; interferon;
KW   interleukin; G-protein coupled receptor; thioesterase; inflammation;
KW   multifactorial disease; autoimmune disease; infection;
KW   nervous system disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200147944-A2.
XX
PD   05-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-US35498.
XX
PR   28-DEC-1999;   99US-0173419.
PR   27-DEC-2000; 2000US-0173419.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2001-465210/50.
XX
PT   Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT   oncogenes and histones, useful for diagnosing and treating, e.g.
PT   cancer, autoimmune diseases and infections -
XX
PS   Claim 1; Page 3600; 4143pp; English.
XX
CC   The present invention relates to oligonucleotides encoding polymorphic
CC   variants of proteins related to amylases, amyloid proteins, angiopoietin,
CC   apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC   histones, kinases, colony stimulating factors, complement related
CC   proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC   G-protein coupled receptors and thioesterases. The present sequence is
CC   one such oligonucleotide. The oligonucleotides and the peptides encoded
CC   by them may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate expression of the proteins listed
CC   above. Disorders that may be prevented, diagnosed and/or treated include
CC   multifactorial diseases with a genetic component, such as autoimmune
CC   diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC   systemic lupus erythromatosus and Grave's disease), inflammation, cancer
CC   (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC   leukaemia), diseases of the nervous system and an infection of pathogenic
CC   organisms.
XX
SQ   Sequence 50 BP; 18 A; 11 C; 13 G; 8 T; 0 other;


alignment_scores:
            Quality:   42.50            Length:     27
              Ratio:  2.833              Gaps:      1
 Percent Similarity:  55.556   Percent Identity: 37.037

alignment_block:
 US-09-836-410A-1 x AAL34450   ..

 Align seg 1/1  to: AAL34450  from: 1  to: 50

      6 LysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValProAr 22
        |||:::||||||:::|||   :::
      3 AAACTATATGAGAAAGCTGGCAAC..................... 26

     22 gLysLeuProLeuAsnPheLeuSerGlyGlu 32
           |||:::||||||||||:::||||||
     27 .......CCAGTGAACTTCCTGGCCGGAGAG 50

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX88379

seq_documentation_block:
ID   AAX88379 standard; DNA; 45 BP.
XX
AC   AAX88379;
XX
DT   30-SEP-1999  (first entry)
XX
DE   Antibody lambda light chain variable region PCR primer Rjlambda0-B.
XX
KW   Haematopoietic; growth factor; PCR primer; mimetic; cell survival;
KW   cell proliferation; cell differentiation; cell activation; agonist;
KW   growth factor inhibitor; nervous system cell; endodermal cell; therapy;
KW   totipotent cell; embryonic stem cell; gene therapy; protection; allergy;
KW   diagnostic; neutropenia; leukemia; aplastic anaemia; thrombocytopenia;
KW   cancer cell; bone marrow transplant; myeloproliferative disease;
KW   antibody; lambda light chain; variable region; ss.
XX
OS   Synthetic.
XX
PN   WO9938008-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-US01331.
XX
PR   23-JAN-1998;   98US-0072253.
XX
PA   (PROL-) PROLIFARON INC.
XX
PI   Bowdish KS;
XX
DR   WPI; 1999-458732/38.
XX
PT   Identification of agonist or inhibitory antibodies to receptors that
PT   control cellular processes, used to modulate, e.g. proliferation of
PT   hematopoietic cells
XX
PS   Example 1; Page 115; 123pp; English.
XX
CC   This invention describes a novel method for identifying agonist or
CC   inhibitory antibodies (Ab) to receptors (R) involved in cell survival,
CC   proliferation, differentiation, or activations. The method is used to
CC   identify Ab that are growth factor mimetics and inhibitors and can
CC   regulate growth, differentiation, survival and activity of many
CC   different cell types, particularly hematopoietic cells (at various
CC   stages, of any lineage), but also nervous system cells, endodermal cells
CC   or totipotent (embryonic stem) cells. Ab are used (i) directly as
CC   therapeutic agents, e.g. for amplifying particular cell types, for ex
CC   vivo proliferation or differentiation of cells for use in gene therapy,
CC   to protect normal cells against chemotherapeutic agents, and to treat
CC   conditions such as allergy, (ii) as diagnostic/research reagents, e.g.
CC   for cell identification and sorting, (iii) to clone receptors and native
CC   factors that they mimic (also potential therapeutic agents). Typical
CC   therapeutic applications are in cases of neutropenia (of any etiology) or
CC   aplastic anemia, bone marrow transplants, myeloproliferative diseases
CC   (e.g. leukemia, thrombocytopenia or allergy) and inhibitory Ab,
CC   optionally coupled to a toxin, are used to kill cancer cells. The method
CC   is not subject to the usual limitations of monoclonal antibody technology
```

```
CC   (limited number of cell fusions) and thus makes possible screening of
CC   very large immune repertoires, nor is it limited by type of regulator or
CC   its origin/site of production. It does not presuppose any biochemical
CC   function, sequence similarity or site of origin, and allows rapid,
CC   inexpensive and simple screening of many samples simultaneously.
CC   AAX88369-X88401 represent PCR primers used in the method of the
CC   invention.
XX
SQ   Sequence 45 BP; 10 A; 13 C; 15 G; 7 T; 0 other;


alignment_scores:
            Quality:   42.00              Length:     14
              Ratio:   3.818                Gaps:      0
 Percent Similarity:   78.571   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AAX88379   ..

 Align seg 1/1  to: AAX88379  from: 1  to: 45

      91 AspAspGlyLysGluGluProProThrThrLeuLeuTrpVal 104
         :::|||   :::|||   |||||||:::|||:::   ||||||
       2 GAAGATCTAGAGGAACCACCGCCTGTGACGGTCAGCTGGGTC 43

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX23741

seq_documentation_block:
ID   AAX23741 standard; DNA; 45 BP.
XX
AC   AAX23741;
XX
DT   28-JUN-1999  (first entry)
XX
DE   Human tumour xenograft scFv library V1 primer RJlambda0-B.
XX
KW   Tumour homing peptide; human; tumour; diagnosis; endothelial cell;
KW   angiogenic vasculature; anti-tumour; anti-inflammatory; anti-angiogenic;
KW   anti-arthritic; NGR receptor; inhibitor; angiogenesis; anticancer drug;
KW   prognosis; inflammation; regeneration; wounded tissue; targetting;
KW   macular degeneration; diabetic retinopathy; rheumatoid arthritis;
KW   occlusive thrombus; xenograft; PCR; primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9913329-A1.
XX
PD   18-MAR-1999.
XX
PF   08-SEP-1998;   98WO-US18895.
XX
PR   25-AUG-1998;   98US-0139802.
PR   10-SEP-1997;   97US-0926914.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Pasqualini R,  Ruoslahti E;
XX
DR   WPI; 1999-215158/18.
XX
PT   Identifying molecules that home to angiogenic vasculature used as
PT   targets for anticancer agents
XX
PS   Example X; Page 136; 180pp; English.
XX
CC   This invention describes novel peptides which home to angiogenic
CC   vasculature, specifically of a tumour and which have anti-tumour,
CC   anti-inflammatory, anti-angiogenic and anti-arthritic activity. Such
CC   molecules are identified by treating a purified NGR receptor with a test
CC   compound and identifying compounds that bind specifically to the NGR
CC   receptor. The peptides of the invention are inhibitors of angiogenesis
CC   and can be used to produce conjugates for delivering agents to
CC   angiogenic vasculature, particularly anticancer drugs or an imaging
CC   agent, for diagnosis or prognosis. These conjugates may be directed to
CC   non-tumour angiogenic vasculature, e.g. that present in inflammatory,
CC   regenerating or wounded tissue, e.g. for treatment of macular
CC   degeneration, diabetic retinopathy or rheumatoid arthritis. The peptides
CC   provide specific targeting to tumours, especially their supporting
CC   vasculature, since the NGR receptor is exposed to the circulation only in
CC   angiogenic vasculature. Precise targeting should reduce the systemic
CC   toxicity of anticancer drugs in the conjugates. Complete killing of all
CC   target cells may not be essential since partial denudation of endothelium
CC   may result in an occlusive thrombus, and endothelial cells are unlikely
CC   to become resistant to anticancer agents nor to lose the targeting
CC   receptor. AAX23730-X23747 and AAX23883 are primers used in the method of
CC   the invention.
XX
SQ   Sequence 45 BP; 10 A; 13 C; 15 G; 7 T; 0 other;


alignment_scores:
            Quality:   42.00              Length:     14
              Ratio:   3.818                Gaps:      0
 Percent Similarity:   78.571   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AAX23741   ..

 Align seg 1/1  to: AAX23741  from: 1  to: 45

      91 AspAspGlyLysGluGluProProThrThrLeuLeuTrpVal 104
         :::|||   :::|||   |||||||:::|||:::   ||||||
       2 GAAGATCTAGAGGAACCACCGCCTGTGACGGTCAGCTGGGTC 43

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ67157

seq_documentation_block:
ID   AAZ67157 standard; DNA; 47 BP.
XX
AC   AAZ67157;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:1504.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,G)
FT                   /*tag= a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
```

```
PS   Claim 1; Page 551; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 19 A; 5 C; 10 G; 13 T; 0 other;


alignment_scores:
            Quality:   42.00            Length:     10
              Ratio:   5.250              Gaps:      0
 Percent Similarity:   80.000   Percent Identity: 70.000

alignment_block:
 US-09-836-410A-1 x AAZ67157   ..

 Align seg 1/1  to: AAZ67157  from: 1  to: 47

       6 LysIleTyrGluGluAlaTrpThrLysTyr 15
         ||||||||||||      |||:::||||||
      17 AAAATATATGAGACAGAGTGGGCCAAATAT 46

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH75873

seq_documentation_block:
ID   AAH75873 standard; DNA; 41 BP.
XX
AC   AAH75873;
XX
DT   26-OCT-2001  (first entry)
XX
DE   Human reverse transcriptase 13 coding sequence probe #1.
XX
KW   Human; reverse transcriptase 13; cytostatic; virucide; immunomodulatory;
KW   antiinflammatory; haemostatic; gene therapy; malignant tumour;
KW   haemopathy; HIV infection; immunological disease; inflammation;
KW   developmental disorder; probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164893-A1.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-CN00280.
XX
PR   02-MAR-2000; 2000CN-0111806.
XX
PA   (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2001-550183/61.
XX
PT   New human reverse transcriptase 13 for diagnosing and treating
PT   developmental disorders, malignant tumor, hemopathy, human
PT   immunodeficiency virus infection, immunological diseases and
PT   inflammations  -
XX
PS   Example 7; Page 15; 34pp; Chinese.
XX
CC   The present invention relates to human reverse transcriptase 13 and its
CC   coding sequence (see AAH75868 and AAG66428). The reverse transcriptase
CC   and its coding sequence are useful in the diagnosis and treatment of
CC   malignant tumour, haemopathy, HIV infection, immunological diseases,
CC   various inflammations and developmental disorders. The present sequence
CC   is a probe, which was used in an example from the present invention.
XX
SQ   Sequence 41 BP; 18 A; 7 C; 6 G; 10 T; 0 other;


alignment_scores:
            Quality:   41.00            Length:     14
              Ratio:   3.417              Gaps:      1
 Percent Similarity:   85.714   Percent Identity: 42.857

alignment_block:
 US-09-836-410A-1 x AAH75873/rev   ..

 Align seg 1/1  to reverse of: AAH75873  from: 1  to: 41

     435 TrpLeuHisGluCysMetIleArgLeuPheHisSerValCys 448
         |||||||||:::|||:::      |||:::::::::::|||
      39 TGGCTGCATAAATGTCTT......CTTTTGAGAAATATCTGT 4

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX05908

seq_documentation_block:
ID   AAX05908 standard; DNA; 47 BP.
XX
AC   AAX05908;
XX
DT   07-MAY-1999  (first entry)
XX
DE   Oligonuleotide probe mecA945-29A18P.
XX
KW   Hybridization; RNase H; scissile linkage; nucleic acid detection;
KW   gene detection; polyamine; probe; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        14..17
FT                   /*tag=  a
XX
PN   WO9901570-A2.
XX
PD   14-JAN-1999.
XX
PF   03-JUL-1998;   98WO-CA00631.
XX
PR   22-JUN-1998;   98US-0090273.
PR   03-JUL-1997;   97US-0051827.
PR   18-MAY-1998;   98US-0086021.
XX
PA   (IDBI-) ID BIOMEDICAL CORP.
XX
PI   Bryan RN,  Cloney LP,  Farnworth BA,  Marostenmaki AJ;
XX
DR   WPI; 1999-106070/09.
XX
PT   Increasing the hybridization rate between two nucleic acids - using
PT   ribonuclease H (RNase H) and/or a polyamine, useful for detecting
PT   nucleic acids of interest in a sample
XX
PS   Example 1; Page 15; 45pp; English.
XX
CC   The invention relates to methods of increasing the hybridization rate
CC   between two nucleic acids. One method comprises construction of two
CC   nucleic acids and a polyamine, and hybridizing both nucleic acids
CC   together, under suitable conditions. Also provided is a similar method
CC   involving two nucleic acids with RNase H, where both nucleic acids do
```

```
CC   not contain a scissile linkage, and if one nucleic acid is DNA, then the
CC   other is RNA. The methods are useful for (in)directly detecting nucleic
CC   acids of interest in a sample. They may be applied to short and long
CC   nucleic acids. The acceleration of rate reactions in gene detection
CC   assays, using RNase H and/or polyamines (under conditions of low salt
CC   concentration), produces a lower background and, therefore, a greater
CC   signal to noise ratio. Sequences AAX05903-914 represent oligonucleotide
CC   probes used in the course of the invention.
XX
SQ   Sequence 47 BP; 37 A; 1 C; 7 G; 2 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:     15
              Ratio:  3.727                Gaps:      0
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-836-410A-1 x AAX05908   ..

 Align seg 1/1  to: AAX05908  from: 1  to: 47

     342 AsnAlaGluLysGluLysProGlnArgAsnProLysLysLysLys 356
         |||   ||||||||||||      ::::::   ||||||||||||
       1 AATAGAGAAAAAGAAAAAAGATGGCAAAGAAAAAAAAAAAAAAAA 45

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX15480

seq_documentation_block:
ID   AAX15480 standard; DNA; 47 BP.
XX
AC   AAX15480;
XX
DT   07-MAY-1999  (first entry)
XX
DE   Probe mecA945-29A18P for detecting an antibiotic resistant mecA gene.
XX
KW   Antibiotic resistant mecA gene; transmission; treatment;
KW   methicillin resistant; Staphylococcus; DNA/RNA hybrid; probe; ss.
XX
OS   Synthetic.
OS   Staphylococcus sp.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        14..17
FT                   /*tag=  a
XX
PN   WO9901572-A2.
XX
PD   14-JAN-1999.
XX
PF   03-JUL-1998;   98WO-CA00633.
XX
PR   22-JUN-1998;   98US-0090276.
PR   03-JUL-1997;   97US-0051643.
PR   18-MAY-1998;   98US-0086020.
XX
PA   (IDBI-) ID BIOMEDICAL CORP.
XX
PI   Bekkaoui F,  Cloney LP;
XX
DR   WPI; 1999-106072/09.
XX
PT   Method for determining the presence of an antibiotic resistant mecA
PT   gene in a sample - using a scissile link containing nucleic acid
PT   probe for antibiotic resistant mecA gene
XX
PS   Example 1; Page 16; 59pp; English.
XX
CC   The present sequence represents a probe used for determining the
CC   presence of an antibiotic resistant mecA gene in a biological sample.
CC   The method provides a means for the rapid detection, for both
CC   the prevention of transmission and treatment of, methicillin resistant
CC   Staphylococcus species.
XX
SQ   Sequence 47 BP; 37 A; 1 C; 7 G; 2 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:     15
              Ratio:  3.727                Gaps:      0
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-836-410A-1 x AAX15480   ..

 Align seg 1/1  to: AAX15480  from: 1  to: 47

     342 AsnAlaGluLysGluLysProGlnArgAsnProLysLysLysLys 356
         |||   ||||||||||||      ::::::   ||||||||||||
       1 AATAGAGAAAAAGAAAAAAGATGGCAAAGAAAAAAAAAAAAAAAA 45

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ67285

seq_documentation_block:
ID   AAZ67285 standard; DNA; 47 BP.
XX
AC   AAZ67285;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:1632.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,C)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 578; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
```

```
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 20 A; 6 C; 10 G; 11 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:     15
              Ratio:  4.556                Gaps:      0
 Percent Similarity:  60.000   Percent Identity: 46.667

alignment_block:
 US-09-836-410A-1 x AAZ67285/rev   ..

 Align seg 1/1  to reverse of: AAZ67285  from: 1  to: 47

      97 ProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHis 111
         |||   |||   :::||||||:::   |||   |||      |||
      47 CCGACAACCTTTATGCTGTGGATTTTTTACATTTTACTTTTACAC 3

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:AAN92016

seq_documentation_block:
ID   AAN92016 standard; DNA; 50 BP.
XX
AC   AAN92016;
XX
DT   17-APR-1990 (first entry)
XX
DE   Sequence probe complementary to Neisseria gonorrhoeae genomic sequence
DE   SSJK1 combined with the LLA2C amplifier sequence.
XX
KW   Neisseria gonorrhoeae genomic sequence SSJK1; LLA2C amplifier sequence;
KW   file 'rcjk'; jkl.probes15(50).
XX
OS   Neisseria gonorrhoeae.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..30
FT                   /*tag=  a
FT                   /"sequence probe"
FT   misc_feature    31..50
FT                   /*tag=  b
FT                   /"LLA2C amplifier sequence"
XX
PN   WO8903891-A.
XX
PD   05-MAY-1989.
XX
PF   14-OCT-1988;   88WO-US03644.
XX
PR   30-SEP-1988;   88US-0252638, US-109282.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Urdea MS,  Warner B,  Running JA,  Kolberg JA,  Clyne JM;
PI   Sanchez-Pescador R;
XX
DR   WPI; 1989-150787/20.
XX
PT   Nucleic acid multimer for hybridisation assays
PT   - having single-stranded oligo-nucleotide units
PT   capable of binding specifically to sequences of interest.
XX
PS   Fig 14; ; 112pp; English.
XX
CC   The sequence probe (tag a ) is complementary to N. gonorrhoeae genomic
CC   sequence SSJK1 from the file 'rcjk'. It is used to assay crude cellular
CC   lysates and genomic DNA from different bacteria. It is called
CC   jkl.probes15(50).
XX
SQ   Sequence 50 BP; 11 A; 12 C; 13 G; 14 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:     13
              Ratio:  4.100                Gaps:      0
 Percent Similarity:  76.923   Percent Identity: 53.846

alignment_block:
 US-09-836-410A-1 x AAN92016/rev   ..

 Align seg 1/1  to reverse of: AAN92016  from: 1  to: 50

     519 GlySerLeuThrAsnArgAsnLeuGlnThrCysMetGlu 531
         |||   :::   |||:::|||||||:::   |||||||||
      44 GGTCCTATGCCTAATCAGAATCTGCATATCTGCATGGAG 6

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL34455

seq_documentation_block:
ID   AAL34455 standard; DNA; 50 BP.
XX
AC   AAL34455;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Human SNP oligonucleotide #7663.
XX
KW   Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
KW   neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
KW   amyloid protein; angiopoietin; apoptosis related protein; cadherin;
KW   cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
KW   complement related protein; cytochrome; kinesin; cytokine; interferon;
KW   interleukin; G-protein coupled receptor; thioesterase; inflammation;
KW   multifactorial disease; autoimmune disease; infection;
KW   nervous system disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200147944-A2.
XX
PD   05-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-US35498.
XX
PR   28-DEC-1999;   99US-0173419.
PR   27-DEC-2000; 2000US-0173419.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2001-465210/50.
XX
PT   Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT   oncogenes and histones, useful for diagnosing and treating, e.g.
PT   cancer, autoimmune diseases and infections -
XX
PS   Claim 1; Page 3601; 4143pp; English.
XX
CC   The present invention relates to oligonucleotides encoding polymorphic
CC   variants of proteins related to amylases, amyloid proteins, angiopoietin,
CC   apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC   histones, kinases, colony stimulating factors, complement related
CC   proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC   G-protein coupled receptors and thioesterases. The present sequence is
CC   one such oligonucleotide. The oligonucleotides and the peptides encoded
CC   by them may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate expression of the proteins listed
CC   above. Disorders that may be prevented, diagnosed and/or treated include
CC   multifactorial diseases with a genetic component, such as autoimmune
CC   diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
```

CC systemic lupus erythromatosus and Grave's disease), inflammation, cancer
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC leukaemia), diseases of the nervous system and an infection of pathogenic
CC organisms.
XX
SQ Sequence 50 BP; 10 A; 13 C; 22 G; 5 T; 0 other;

alignment_scores:
Quality: 41.00 Length: 12
Ratio: 4.100 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:
US-09-836-410A-1 x AAL34455 ..

Align seg 1/1 to: AAL34455 from: 1 to: 50

```
364 GlyGlyProLysGluGluLeuIleProGluLysLeu 375
    |||||||||   |||||||||:::   |||:::|||
 14 GGAGGGCCGCCAGAGGAGCTGCTGCGGGAGAACCTA 49
```

OM of: US-09-836-410A-1 to: EST:* out_format : pfs

Date: Jul 20, 2002 4:12 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_115032_29691/app_query.fasta_1.660
-DB=EST -QFMT=fastap -SUFFIX=p2n15to50.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=15 -MAXLEN=50
-USER=US09836410_@CGN1_1_2960 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1532.560000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_gss:AZ774479	-	48.00	92.20	6.9e+04	33	! AZ774479 2M0004A05F Mouse 10kb p
gb_est1:AI657570	-	47.50	89.72	9.5e+04	40	! AI657570 fc15c02.y1 Zebrafish Wa
gb_est1:AI223998	+	47.00	87.71	1.2e+05	46	! AI223998 qx12h07.x1 NCI_CGAP_Lym
gb_est2:BF054800	+	45.00	83.94	2.0e+05	50	! BF054800 7i71d09.y1 NCI_CGAP_Brn
gb_est1:AA590944	+	43.00	80.92	2.9e+05	50	! AA590944 vm25f02.r1 Knowles Solt
gb_est1:AI813747	+	42.00	82.13	2.5e+05	37	! AI813747 wk79a02.x1 NCI_CGAP_Par
gb_est2:BM393347	-	41.00	78.09	4.2e+05	49	! BM393347 50071-2-9-B10.f.1 Child
gb_est2:BM395447	-	41.00	78.09	4.2e+05	49	! BM395447 50072-2-9-B10.f.1 Child
gb_est1:AU104189	+	41.00	77.91	4.3e+05	50	! AU104189 AU104189 Sugano Homo sa
gb_est1:AU107491	+	41.00	77.91	4.3e+05	50	! AU107491 AU107491 Sugano Homo sa
gb_gss:AZ484607	-	40.50	78.10	4.2e+05	45	! AZ484607 1M0311N10F Mouse 10kb p
gb_gss:AZ860401	+	40.00	77.76	4.3e+05	43	! AZ860401 2M0166E22F Mouse 10kb p
gb_est1:AL587842	-	40.00	77.55	4.5e+05	44	! AL587842 AL587842 BP Chicken Bra
gb_est1:AU103783	-	40.00	76.40	5.2e+05	50	! AU103783 AU103783 Sugano Homo sa
gb_est1:AU104282	+	40.00	76.40	5.2e+05	50	! AU104282 AU104282 Sugano Homo sa
gb_est1:AU107535	+	40.00	76.40	5.2e+05	50	! AU107535 AU107535 Sugano Homo sa
gb_est1:AA574989	+	40.00	76.40	5.2e+05	50	! AA574989 vm34a03.r1 Knowles Solt
gb_gss:AZ582769	-	39.50	76.20	5.3e+05	47	! AZ582769 1M0376B22F Mouse 10kb p
gb_est1:AI500599	+	39.50	75.83	5.6e+05	49	! AI500599 tn93d07.x1 NCI_CGAP_Ut2
gb_est2:BJ043594	-	39.50	75.83	5.6e+05	49	! BJ043594 BJ043594 NIBB Mochii no
gb_est2:BI330882	+	39.00	75.84	5.5e+05	45	! BI330882 602981270F1 NCI_CGAP_Li
gb_est1:AL627881	-	39.00	75.45	5.8e+05	47	! AL627881 AL627881 XGC-gastrula S
gb_est1:AA254893	+	39.00	75.07	6.1e+05	49	! AA254893 mz78d07.r1 Soares mouse
gb_est1:AA966391	+	39.00	74.89	6.3e+05	50	! AA966391 w4f01a1.r1 Aspergillus
gb_est1:AU107471	+	39.00	74.89	6.3e+05	50	! AU107471 AU107471 Sugano Homo sa
gb_est1:AU107473	+	39.00	74.89	6.3e+05	50	! AU107473 AU107473 Sugano Homo sa
gb_est1:AU107474	+	39.00	74.89	6.3e+05	50	! AU107474 AU107474 Sugano Homo sa
gb_est1:AU107475	+	39.00	74.89	6.3e+05	50	! AU107475 AU107475 Sugano Homo sa
gb_est1:AU107481	+	39.00	74.89	6.3e+05	50	! AU107481 AU107481 Sugano Homo sa
gb_est1:AU107483	+	39.00	74.89	6.3e+05	50	! AU107483 AU107483 Sugano Homo sa
gb_est1:AU107486	+	39.00	74.89	6.3e+05	50	! AU107486 AU107486 Sugano Homo sa
gb_est1:AU107487	+	39.00	74.89	6.3e+05	50	! AU107487 AU107487 Sugano Homo sa
gb_est1:AU107488	+	39.00	74.89	6.3e+05	50	! AU107488 AU107488 Sugano Homo sa
gb_est1:AU107490	+	39.00	74.89	6.3e+05	50	! AU107490 AU107490 Sugano Homo sa
gb_est1:AU107492	+	39.00	74.89	6.3e+05	50	! AU107492 AU107492 Sugano Homo sa
gb_est1:AU107493	+	39.00	74.89	6.3e+05	50	! AU107493 AU107493 Sugano Homo sa
gb_est1:AU107494	+	39.00	74.89	6.3e+05	50	! AU107494 AU107494 Sugano Homo sa
gb_est1:AU107495	+	39.00	74.89	6.3e+05	50	! AU107495 AU107495 Sugano Homo sa
gb_est1:AU107496	+	39.00	74.89	6.3e+05	50	! AU107496 AU107496 Sugano Homo sa
gb_est1:AU107497	+	39.00	74.89	6.3e+05	50	! AU107497 AU107497 Sugano Homo sa
gb_est1:AU107499	+	39.00	74.89	6.3e+05	50	! AU107499 AU107499 Sugano Homo
gb_est1:AU107500	+	39.00	74.89	6.3e+05	50	! AU107500 AU107500 Sugano Homo
gb_est1:AU107501	+	39.00	74.89	6.3e+05	50	! AU107501 AU107501 Sugano Homo
gb_est1:AU107502	+	39.00	74.89	6.3e+05	50	! AU107502 AU107502 Sugano Homo
gb_est1:AU107503	+	39.00	74.89	6.3e+05	50	! AU107503 AU107503 Sugano Homo

seq_name: gb_gss:AZ774479

seq_documentation_block:
LOCUS AZ774479 33 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0004A05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0004A05 F, DNA sequence.
ACCESSION AZ774479
VERSION AZ774479.1 GI:12899972
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 33)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0004 row: A column: 05
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 33.
FEATURES Location/Qualifiers
source 1. .33
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0004A05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 0 a 0 c 6 g 27 t
ORIGIN

alignment_scores:

```
          Quality:   48.00              Length:      10
            Ratio:  4.800                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 80.000

alignment_block:
 US-09-836-410A-1 x AZ774479/rev    ..

 Align seg 1/1  to reverse of: AZ774479  from: 1  to: 33

     347 LysProGlnArgAsnProLysLysLysLys 356
         |||||:::::|||||||||||||||||||
      32 AAACCAAAAAAAAACCCCAAAAAAAAAAA 3

seq_name: gb_est1:AI657570

seq_documentation_block:
LOCUS       AI657570                40 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fc15c02.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3721442 5' similar to TR:O13017 O13017 WINGED HELIX
            TRANSCRIPTIONAL FACTOR MFH-1. ;, mRNA sequence.
ACCESSION   AI657570
VERSION     AI657570.1  GI:4755238
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3721442"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT        8 a    11 c     14 g      7 t
ORIGIN

alignment_scores:
          Quality:   47.50              Length:      11
            Ratio:  4.750                Gaps:       1
 Percent Similarity:  90.909   Percent Identity: 90.909

alignment_block:
 US-09-836-410A-1 x AI657570/rev    ..

 Align seg 1/1  to reverse of: AI657570  from: 1  to: 40

     560 ProTyrAlaLeuAlaPheMetProProGlyTyr 570
         |||||||||||||||   |||||||||||||||
      33 CCATATGCGCTAGCC...ATGCCGCCTGGGTAC 4

seq_name: gb_est1:AI223998

seq_documentation_block:
LOCUS       AI223998                46 bp    mRNA    linear   EST 21-DEC-1998
DEFINITION  qx12h07.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2001181 3'
            similar to TR:Q04154 Q04154 SALIVARY PROLINE-RICH PROTEIN RP15
            PRECURSOR. ;contains element MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI223998
VERSION     AI223998.1  GI:3806711
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Trace considered overall poor quality
            Insert Length: 1534   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2001181"
                     /clone_lib="NCI_CGAP_Lym12"
                     /tissue_type="lymphoma, follicular mixed small and large
                     cell"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.25 kb. Life Technologies
                     catalog #: 11547-015"
BASE COUNT       19 a     17 c     10 g      0 t
ORIGIN

alignment_scores:
```

```
           Quality:   47.00              Length:     19
             Ratio:  4.273                Gaps:      1
 Percent Similarity:  57.895   Percent Identity: 52.632

alignment_block:
 US-09-836-410A-1 x AI223998   ..

 Align seg 1/1  to: AI223998  from: 1  to: 46

     348 ProGlnArgAsnProLysLysLysLysAspAspAspAspGluGluIleGl 364
         ||||||:::|||||||||||||||                        ||
      14 CCCCAAAAAAACCCCAAAAAAAAA.....................GG 39

     364 yGlyPro 366
         |||||||
      40 GGGGCCC 46

seq_name: gb_est2:BF054800

seq_documentation_block:
LOCUS       BF054800                 50 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  7i71d09.y1 NCI_CGAP_Brn20 Homo sapiens cDNA clone IMAGE:3340145 5',
            mRNA sequence.
ACCESSION   BF054800
VERSION     BF054800.1  GI:10808696
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: 7i71d09.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Robert Jenkins, M.D., Mark Israel, M.D., Jim
            Jacobson, Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3340145"
                     /clone_lib="NCI_CGAP_Brn20"
                     /tissue_type="oligodendroglioma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pAMP1; mRNA made from
                     oligodendroglioma tissue, cDNA made by oligo-dT  priming.
                     Directionally cloned.  Size-selected on agarose  gel,
                     average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       33 a      3 c     11 g      3 t
ORIGIN


alignment_scores:
           Quality:   45.00              Length:     13
             Ratio:  4.091                Gaps:      0
 Percent Similarity:  84.615   Percent Identity: 69.231

alignment_block:
 US-09-836-410A-1 x BF054800   ..

 Align seg 1/1  to: BF054800  from: 1  to: 50

     344 GluLysGluLysProGlnArgAsnProLysLysLysLys 356
         ||||||||||||   :::|||:::   ||||||||||||
       2 GAAAAAGAAAAGAAGGAAAGAAGGACAAAAAAAAAAAAA 40

seq_name: gb_est1:AA590944

seq_documentation_block:
LOCUS       AA590944                 50 bp    mRNA    linear   EST 16-SEP-1997
DEFINITION  vm25f02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:991227 5', mRNA sequence.
ACCESSION   AA590944
VERSION     AA590944.1  GI:2404257
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:563507.
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:991227"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.   Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                     inserts) and B3."
BASE COUNT       47 a      1 c      1 g      1 t
ORIGIN


alignment_scores:
           Quality:   43.00              Length:     16
             Ratio:  3.308                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AA590944   ..

 Align seg 1/1  to: AA590944  from: 1  to: 50

     341 LysAsnAlaGluLysGluLysProGlnArgAsnProLysLysLysLys 356
         ||||||   :::|||:::|||   :::::::::   ||||||||||||
       1 AAGAATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 48
```

```
seq_name: gb_est1:AI813747

seq_documentation_block:
LOCUS       AI813747                37 bp    mRNA    linear   EST 07-MAR-2000
DEFINITION  wk79a02.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2421578 3'
            similar to TR:Q39599 Q39599 EXTENSIN. ;contains element MSR1
            repetitive element ;, mRNA sequence.
ACCESSION   AI813747
VERSION     AI813747.1  GI:5424962
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 718   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2421578"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       19 a     16 c      1 g      1 t
ORIGIN


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  4.667               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-836-410A-1 x AI813747   ..

 Align seg 1/1  to: AI813747  from: 1  to: 37

     348 ProGlnArgAsnProLysLysLysLys 356
         ||||||:::::|||||||||||||||||
       6 CCCCAAAAAAGCCCCAAAAAAAAAAAA 32

seq_name: gb_est2:BM393347

seq_documentation_block:
LOCUS       BM393347                49 bp    mRNA    linear   EST 17-JAN-2002
DEFINITION  50071-2-9-B10.f.1 Chilcoat/Turkewitz cDNA (small fraction)
            Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION   BM393347
VERSION     BM393347.1  GI:18193400
KEYWORDS    EST.
SOURCE      Tetrahymena thermophila.
  ORGANISM  Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
            ,J. and Klobutcher,L.
  TITLE     EST from Tetrahymena thermophila, strain CU428.1, growing cells
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Turkewitz AP
            Molecular Genetics and Cell Biology
            University of Chicago
            920 E. 58th Street, Chicago, IL 60637, USA
            Tel: 773 702 4374
            Fax: 773 702 3172
            Email: apturkew@midway.uchicago.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Tetrahymena thermophila"
                     /strain="CU428.1"
                     /db_xref="taxon:5911"
                     /clone_lib="Chilcoat/Turkewitz cDNA (small fraction)"
                     /note="Vector: BlueScript2 SK+; Details on library
                     preparation can be found in Chilcoat and Turkewitz (2001)
                     Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT       19 a      9 c      7 g     13 t      1 others
ORIGIN


alignment_scores:
            Quality:   41.00             Length:     10
              Ratio:  4.556               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x BM393347/rev   ..

 Align seg 1/1  to reverse of: BM393347  from: 1  to: 49

     236 CysHisGluIleGluArgHisPheIleGlu 245
         ||||||:::::|||   |||||||:::::
      47 TGCCACCGGGTAGAGTGGCATTTTGTTCAA 18

seq_name: gb_est2:BM395447

seq_documentation_block:
LOCUS       BM395447                49 bp    mRNA    linear   EST 17-JAN-2002
DEFINITION  50072-2-9-B10.f.1 Chilcoat/Turkewitz cDNA (large fraction)
            Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION   BM395447
VERSION     BM395447.1  GI:18195500
KEYWORDS    EST.
SOURCE      Tetrahymena thermophila.
  ORGANISM  Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
            ,J. and Klobutcher,L.
  TITLE     EST from Tetrahymena thermophila, strain CU428.1, growing cells
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Turkewitz AP
            Molecular Genetics and Cell Biology
            University of Chicago
            920 E. 58th Street, Chicago, IL 60637, USA
            Tel: 773 702 4374
            Fax: 773 702 3172
            Email: apturkew@midway.uchicago.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Tetrahymena thermophila"
                     /strain="CU428.1"
                     /db_xref="taxon:5911"
                     /clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
```

```
                /note="Vector: BlueScript2 SK+; Details on library
                preparation can be found in Chilcoat and Turkewitz (2001)
                Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT       19 a      9 c      7 g     13 t      1 others
ORIGIN


alignment_scores:
           Quality:   41.00             Length:     10
             Ratio:  4.556               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x BM395447/rev   ..

 Align seg 1/1  to reverse of: BM395447  from: 1  to: 49

     236 CysHisGluIleGluArgHisPheIleGlu 245
         ||||||::::::|||   ||||||::::::
      47 TGCCACCGGGTAGAGTGGCATTTTGTTCAA 18

seq_name: gb_est1:AU104189

seq_documentation_block:
LOCUS       AU104189                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU104189 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            LNG04978, mRNA sequence.
ACCESSION   AU104189
VERSION     AU104189.1  GI:13553710
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="LNG04978"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       19 a     15 c      5 g     11 t
ORIGIN


alignment_scores:
           Quality:   41.00             Length:     12
             Ratio:  3.727               Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 41.667

alignment_block:
 US-09-836-410A-1 x AU104189   ..

 Align seg 1/1  to: AU104189  from: 1  to: 50

     550 AlaTyrArgAlaSerCysHisLysLeuProTyr 561
         ::::::::::   |||||||||:::::::||||||:::
      15 TCGTTCCACAGAAGCTGCCATCAAGTATTTCCTCAC 50

seq_name: gb_est1:AU107491

seq_documentation_block:
LOCUS       AU107491                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU107491 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP05672, mRNA sequence.
ACCESSION   AU107491
VERSION     AU107491.1  GI:13557012
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP05672"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT        4 a     18 c     16 g     12 t
ORIGIN


alignment_scores:
           Quality:   41.00             Length:     15
             Ratio:  3.417               Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 53.333

alignment_block:
 US-09-836-410A-1 x AU107491   ..

 Align seg 1/1  to: AU107491  from: 1  to: 50

     420 LeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisPro 434
         |||:::|||:::      :::|||||||:::   ||||||||||||
       5 TTGCGCAGTACTGCTAGCGGCTTCGCGGTTCGGTCCTCGCACCCG 49

seq_name: gb_gss:AZ484607

seq_documentation_block:
LOCUS       AZ484607                 45 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0311N10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0311N10 F, DNA sequence.
ACCESSION   AZ484607
VERSION     AZ484607.1  GI:10649619
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
```

```
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0311  row: N  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 45.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0311N10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       11 a     13 c     12 g      9 t
ORIGIN


alignment_scores:
            Quality:   40.50            Length:     20
              Ratio:  4.050              Gaps:      1
 Percent Similarity:  50.000   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AZ484607/rev   ..

 Align seg 1/1  to reverse of: AZ484607  from: 1  to: 45

      12 TrpThrLysTyrProArgGlyLeuValProArgLysLeuProLeuAsnPh 28
         |||         ||||||||||||||||||
      40 TGGCCACTGATCCCCAGGGGTCTGGTTCCC..................... 11

      28 eLeuSerGly 31
          |||||||||
      10 .CTAAGTGGA 2

seq_name: gb_gss:AZ860401

seq_documentation_block:
LOCUS       AZ860401                 43 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0166E22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0166E22 F, DNA sequence.
ACCESSION   AZ860401
VERSION     AZ860401.1  GI:13055685
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0166  row: E  column: 22
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 43.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0166E22"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       27 a     15 c      1 g      0 t
ORIGIN


alignment_scores:
            Quality:   40.00            Length:     14
              Ratio:  4.000              Gaps:      0
 Percent Similarity:  71.429   Percent Identity: 50.000

alignment_block:
 US-09-836-410A-1 x AZ860401   ..

 Align seg 1/1  to: AZ860401  from: 1  to: 43

     341 LysAsnAlaGluLysGluLysProGlnArgAsnProLysLys 354
         ||||||         :::::|||   :::|||||||||||||
       2 AAAAACCAACCAACCAACCAACCAACCAAAAACCCGAAAAAA 43

seq_name: gb_est1:AL587842
```

```
seq_documentation_block:
LOCUS       AL587842                 44 bp    mRNA    linear   EST 02-MAR-2001
DEFINITION  AL587842 BP Chicken Brain Library Gallus gallus cDNA clone
            ROS064C05, mRNA sequence.
ACCESSION   AL587842
VERSION     AL587842.1  GI:13192876
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Murray,F.
  TITLE     BP Chicken Brain Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Frazer Murray
            Dept. Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UK
            Tel: +44 (0)131 527 4200
            Fax: +44 (0)131 440 0434
            Email: frazer.murray@bbsrc.ac.uk
            GCGGCCGCTTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from Clonetech
            (*6854-
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /clone="ROS064C05"
                     /clone_lib="BP Chicken Brain Library"
                     /tissue_type="Brain"
                     /dev_stage="Unknown"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Not1; Site_2: Sal1; Cloned
                     unidirectionally. Primer: Oligo dT. 5' adaptor sequence:
                     5' TCGACCTCGAG 3' ; 3' adaptor sequence: 5'
                     GCGGCCGCTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from
                     Clonetech (*6854-1)"
BASE COUNT        1 a      0 c      4 g     39 t
ORIGIN


alignment_scores:
            Quality:   40.00             Length:     12
              Ratio:  3.636               Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 58.333

alignment_block:
 US-09-836-410A-1 x AL587842/rev    ..

 Align seg 1/1  to reverse of: AL587842  from: 1  to: 44

     345 LysGluLysProGlnArgAsnProLysLysLysLys 356
         |||:::|||   ::::::::::|||||||||||||||
      42 AAAAAAAAAAAAAAAAAAAAACCAAAAAAAAAAAAA 7

seq_name: gb_est1:AU103783

seq_documentation_block:
LOCUS       AU103783                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU103783 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP14796, mRNA sequence.
ACCESSION   AU103783
VERSION     AU103783.1  GI:13553304
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki;Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP14796"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       10 a     13 c     11 g     16 t
ORIGIN


alignment_scores:
            Quality:   40.00             Length:     15
              Ratio:  4.000               Gaps:      0
 Percent Similarity:  66.667   Percent Identity: 53.333

alignment_block:
 US-09-836-410A-1 x AU103783/rev    ..

 Align seg 1/1  to reverse of: AU103783  from: 1  to: 50

     103 TrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGln 117
         |||:::          |||||||||       ||||||:::||||||
      49 TGGCTACTGCTCACTTTGGCGCAGGGAAGCGACAAACTTGGACAA 5

seq_name: gb_est1:AU104282

seq_documentation_block:
LOCUS       AU104282                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU104282 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP05565, mRNA sequence.
ACCESSION   AU104282
VERSION     AU104282.1  GI:13553803
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP05565"
```

/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT 8 a 13 c 24 g 5 t
ORIGIN

alignment_scores:
Quality: 40.00 Length: 12
Ratio: 4.000 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 50.000

alignment_block:
US-09-836-410A-1 x AU104282 ..

Align seg 1/1 to: AU104282 from: 1 to: 50

```
357 AspAspAspAspGluGluIleGlyGlyProLysGlu 368
    :::|||   |||   ::::::||||||||||:::|||
  7 GAGGACGGCGACACGCGGGTCGGCGGGCCGCAGGAG 42
```

OM of: US-09-836-410A-1 to: N_Geneseq_032802:* out_format : pfs

Date: Jul 20, 2002 3:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_114913_29226/app_query.fasta_1.660
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836410_@CGN1_1_188 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 201.440000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..	
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD22687	+	3104.00	5422.07	1.2e-29			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77156	+	3104.00	5422.07	1.2e-29			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77158	-	3104.00	5422.07	1.2e-29			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS71925	+	2299.50	4011.46	4.3e-2			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77157	-	1768.00	3083.24	2.2e-16			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH16408	+	1747.00	3043.74	3.4e-1			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH16424	+	1521.50	2646.36	4.7e-1			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH81664	-	1176.00	2050.04	7.7e-1			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ15705	+	1104.00	1922.47	9.8e-99			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX99053	+	1104.00	1922.47	9.8e-99			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ15983	+	973.00	1692.45	6.4e-86			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX98777	+	973.00	1692.45	6.4e-86			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS37350	+	680.00	1184.30	1.3e-5			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH12222	-	605.50	1051.29	3.3e-5			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:AAN90541	+	550.50	936.38	8.3e-44			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:AAQ12226	+	549.00	933.66	1.2e-43			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ12696	+	517.00	900.92	7.8e-42			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX98342	+	517.00	900.92	7.8e-42			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH06489	+	472.00	811.51	7.4e-3			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH14477	+	472.00	797.80	4.3e-3			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH81986	-	452.00	788.40	1.4e-3			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS37726	+	420.00	727.15	3.7e-3			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS37727	+	390.00	674.42	3.2e-2			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC21231	+	373.00	647.92	9.6e-28			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX99063	+	319.50	543.70	6.2e-22			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX98340	+	235.00	405.26	3.2e-14			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH99691	-	170.50	258.87	4.5e-0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA70105	+	168.00	255.68	6.8e-06			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH98321	-	165.50	250.15	1.4e-0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT34291	+	164.00	248.24	1.8e-05			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX39679	+	160.50	250.10	1.4e-05			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA83085	+	160.50	250.10	1.4e-0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAC87951	+	160.50	250.10	1.4e-0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH98726	-	152.50	237.28	7.2e-0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ77531	+	150.00	231.16	0.0002			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH23688	+	149.50	218.66	0.0008			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH23695	+	149.50	218.60	0.0008			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH23705	+	149.50	206.30	0.0038			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC48915	+	148.00	231.61	0.0001			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK51987	+	148.00	218.02	0.0009			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAC87952	+	147.50	228.69	0.0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV21209	+	147.00	158.72	1.71			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL04433	+	146.50	224.61	0.0			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL00120	+	145.00	242.75	3.6			
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ94068	+	145.00	222.34	0.00			

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD22687

seq_documentation_block:

```
ID   AAD22687 standard; cDNA; 3418 BP.
XX
AC   AAD22687;
XX
DT   26-FEB-2002  (first entry)
XX
DE   Tubedown-1 (tbdn-1) protein encoding cDNA.
XX
KW   Tubedown-1 protein; tbdn-1; ophthalmological; cytostatic; vulnerary;
KW   cerebroprotective; angiogenesis inhibitor; ocular neovascularisation;
KW   retinal disease; diabetic retinopathy; retinopathy of prematurity;
KW   primary hyperplastic vitreous; macular degeneration; trauma; stroke;
KW   haemorrhagic shock; arthritis; arteriosclerosis; delayed wound healing;
KW   angiofibroma; granulation; nonunion fracture; retrolental fibroplasia;
KW   solid tumour growth; chronic glaucoma; sickle cell retinopathy; cancer;
KW   burn; scar; corneal neovascularisation; rubeosis iritis; uveitis;
KW   gene therapy; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             18..2189
FT                   /*tag= a
FT                   /product= "Tubedown-1 protein #2"
FT   CDS             57..2189
FT                   /*tag= b
FT                   /product= "Tubedown-1 protein #3"
FT   CDS             87..2189
FT                   /*tag= c
FT                   /product= "Tubedown-1 protein #4"
FT   CDS             408..2189
FT                   /*tag= d
FT                   /product= "Tubedown-1 protein #1"
FT   misc_feature    408..2186
FT                   /note= "This region is specifically claimed as
FT                   SEQ ID NO: 6 in claim 5 of the specification"
XX
PN   WO200179506-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12548.
XX
PR   17-APR-2000; 2000US-197977P.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-026032/03.
DR   P-PSDB; AAE13589, AAE13590, AAE13591, AAE13592.
XX
PT   Novel tubedown-1 protein comprising anti-angiogenic activity is useful
PT   for treating angiogenesis-associated disease related to ocular
PT   neovascularization, e.g., diabetic retinopathy, retinopathy of
PT   prematurity -
XX
PS   Claim 10; Page 56-58; 85pp; English.
XX
CC   The present invention relates to tubedown-1 (tbdn-1) proteins and
CC   their corresponding cDNAs. Tbdn-1 proteins having anti-angiogenic
CC   activity are associated with acetyl transferase activity. They
CC   regulate endothelial differentiation through protein acetylation,
CC   DNA-binding or by interacting with and/or acetylating other protein
```

CC targets important for endothelial differentiation. In normal adult
CC eyes, tbdn-1 is expressed highly in the corneal endothelium proper
CC and in the vascular endothelium of the limbus and retina. Tbdn-1
CC proteins are useful for preventing, treating, inhibiting or delaying
CC the onset of angiogenesis-associated disease related to ocular
CC neovascularisation, e.g., a retinal disease, such as preferably
CC diabetic retinopathy or retinopathy of prematurity, or primary
CC hyperplastic vitreous, macular degeneration and any other conditions
CC involving ocular neovascularisation. Tbdn-1 proteins are also useful
CC for treating any pathological neovascularisation condition such as
CC head trauma, spinal trauma, stroke, haemorrhagic shock, arthritis,
CC arteriosclerosis, angiofibroma, delayed wound healing, granulations,
CC cancer, burns, scars, nonunion fractures, retrolental fibroplasia,
CC solid tumour growth. Proteins of the invention are also useful for
CC treating ocular neovascularisation conditions such as chronic glaucoma,
CC sickle cell retinopathy, corneal neovascularisation, rubeosis iritis,
CC uveitis, neovascularisation of the optic nerve. Sequences of the
CC invention are also used in gene therapy. The present sequence is
CC a cDNA encoding tubedown-1 proteins.
XX
SQ Sequence 3418 BP; 1157 A; 604 C; 704 G; 953 T; 0 other;

alignment_scores:
Quality: 3104.00 Length: 593
Ratio: 5.234 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-836-410A-1 x AAD22687 ..

Align seg 1/1 to: AAD22687 from: 1 to: 3418

```
      1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    408 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 457

     17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    458 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTTATCTGGAGAGAAGT 507

     34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    508 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 557

     51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    558 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 607

     67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    608 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 657

     84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    658 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

    101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    708 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 757

    117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    758 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 807

    134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    808 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 857

    151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    858 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 907

    167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    908 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 957

    184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    958 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 1007

    201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1008 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 1057

    217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1058 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 1107

    234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1108 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 1157

    251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1158 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 1207

    267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1208 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1257

    284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1258 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 1307

    301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1308 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 1357

    317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1358 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1407

    334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1408 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1457

    351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1458 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1507

    367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1508 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1557

    384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1558 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1607

    401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1608 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1657

    417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1658 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1707

    434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1708 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1757

    451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1758 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1807

    467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1808 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1857

 484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1858 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1907

 501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1908 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1957

 517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1958 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 2007

 534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2008 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 2057

 551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2058 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 2107

 567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2108 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 2157

 584 laGluThrGluGluLeuAlaAsnGluIle 593
     |||||||||||||||||||||||||||||
2158 CAGAAACGGAAGAACTGGCCAATGAAATC 2186
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77156

seq_documentation_block:

```
ID   AAH77156 standard; cDNA; 3418 BP.
XX
AC   AAH77156;
XX
DT   21-JAN-2002  (first entry)
XX
DE   Human tubedown-1 cDNA.
XX
KW   Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW   bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             408..2189
FT                   /*tag=  a
FT                   /product= "tubedown-1"
XX
PN   WO200179505-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12435.
XX
PR   17-APR-2000; 2000US-197977P.
PR   17-APR-2001; 2001US-0836410.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-017618/02.
DR   P-PSDB; AAG77907.
XX
PT   Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT   overexpression of tubedown-1 protein and are useful to treat
PT   osteosarcoma and Ewing's Sarcoma family of tumours  -
XX
PS   Claim 1; Page 36-38; 56pp; English.
XX
CC   The sequence represents a new human gene, tubedown-1 (tbdn-1). The
CC   invention relates to a novel isolated nucleic acid of the tubedown-1
CC   gene, and antisense nucleic acids to tbdn-1. The polynucleotides and
CC   protein of the invention have cytostatic and osteopathic activity. The
CC   polynucleotides of the invention may be used in antisense-therapy/gene
CC   therapy. They are useful in the treatment of bone tumours, especially
CC   osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC   invention may also be useful for the prevention of metastases from these
CC   types of tumours, either alone or in combination with radiotherapy and/or
CC   chemotherapeutic agents.
XX
SQ   Sequence 3418 BP; 1157 A; 604 C; 704 G; 953 T; 0 other;
```

alignment_scores:

```
           Quality: 3104.00              Length:    593
             Ratio:  5.234                 Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

```
US-09-836-410A-1 x AAH77156   ..

Align seg 1/1  to: AAH77156  from: 1  to: 3418

      1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    408 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 457

     17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    458 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGT 507

     34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    508 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 557

     51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    558 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 607

     67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    608 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 657

     84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    658 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

    101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    708 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 757

    117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    758 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 807

    134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    808 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 857

    151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    858 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 907

    167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    908 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 957

    184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    958 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 1007
```

```
 201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1008 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 1057

 217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1058 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 1107

 234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1108 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 1157

 251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1158 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 1207

 267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1208 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1257

 284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1258 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 1307

 301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1308 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 1357

 317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1358 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1407

 334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1408 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1457

 351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1458 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1507

 367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1508 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1557

 384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1558 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1607

 401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1608 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1657

 417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1658 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1707

 434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1708 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1757

 451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1758 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1807

 467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1808 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1857

 484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1858 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1907

 501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1908 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1957

 517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1958 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 2007

 534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2008 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 2057

 551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2058 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 2107

 567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2108 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 2157

 584 laGluThrGluGluLeuAlaAsnGluIle 593
     ||||||||||||||||||||||||||||
2158 CAGAAACGGAAGAACTGGCCAATGAAATC 2186
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77158

seq_documentation_block:

```
ID   AAH77158 standard; cDNA; 3418 BP.
XX
AC   AAH77158;
XX
DT   21-JAN-2002  (first entry)
XX
DE   Human tubedown-1 base pairs 3418-1 antisense cDNA.
XX
KW   Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW   bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179505-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12435.
XX
PR   17-APR-2000; 2000US-197977P.
PR   17-APR-2001; 2001US-0836410.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-017618/02.
XX
PT   Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT   overexpression of tubedown-1 protein and are useful to treat
PT   osteosarcoma and Ewing's Sarcoma family of tumours  -
XX
PS   Claim 7; Page 39-41; 56pp; English.
XX
CC   The sequence represents tubedown-1 (tbdn-1) bases 3418-1 antisense cDNA.
CC   The invention relates to a novel isolated nucleic acid of the tubedown-1
CC   gene, and antisense nucleic acids to tbdn-1. The polynucleotides and
CC   protein of the invention have cytostatic and osteopathic activity. The
CC   polynucleotides of the invention may be used in antisense-therapy/gene
CC   therapy. They are useful in the treatment of bone tumours, especially
CC   osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC   invention may also be useful for the prevention of metastases from these
CC   types of tumours, either alone or in combination with radiotherapy and/or
CC   chemotherapeutic agents.
XX
SQ   Sequence 3418 BP; 953 A; 704 C; 604 G; 1157 T; 0 other;
```

```
alignment_scores:
           Quality: 3104.00              Length:    593
             Ratio:  5.234                Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-836-410A-1 x AAH77158/rev   ..

Align seg 1/1  to reverse of: AAH77158  from: 1  to: 3418

      1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   3011 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 2962

     17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2961 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTTATCTGGAGAGAAGT 2912

     34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2911 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 2862

     51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2861 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 2812

     67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2811 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 2762

     84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2761 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 2712

    101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2711 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 2662

    117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2661 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 2612

    134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2611 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 2562

    151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2561 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 2512

    167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2511 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 2462

    184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2461 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 2412

    201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2411 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 2362

    217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2361 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 2312

    234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2311 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 2262

    251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2261 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 2212

    267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2211 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 2162

    284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2161 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 2112

    301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2111 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 2062

    317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2061 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 2012

    334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2011 CCCAGATTGAAGAAGAGAAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1962

    351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1961 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1912

    367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1911 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1862

    384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1861 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1812

    401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1811 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1762

    417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1761 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1712

    434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1711 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1662

    451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1661 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1612

    467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1611 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1562

    484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1561 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1512

    501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1511 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1462

    517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1461 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 1412

    534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1411 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 1362

    551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1361 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 1312
```

```
     567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1311 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 1262

     584 laGluThrGluGluLeuAlaAsnGluIle 593
         ||||||||||||||||||||||||||||||
    1261 CAGAAACGGAAGAACTGGCCAATGAAATC 1233

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS71925

seq_documentation_block:
ID   AAS71925 standard; cDNA; 2477 BP.
XX
AC   AAS71925;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7729.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG07738.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 7729; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2477 BP; 859 A; 437 C; 522 G; 659 T; 0 other;

alignment_scores:
            Quality: 2299.50              Length:    608
              Ratio:  4.581                Gaps:     16
 Percent Similarity:  82.566   Percent Identity: 80.099

alignment_block:
 US-09-836-410A-1 x AAS71925   ..

 Align_seg 1/1  to: AAS71925  from: 1  to: 2477

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTGGACTAAATATCCCAG 584

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||:::||||||||||||||||||||||||||||||
     585 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 634

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||:::||||||||||||||||||||||||||||||
     635 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 684

      51 ProProValPheAsnThrLeuArgSerLeu.TyrArgAspLysGluLys. 66
         |||||||||||||||||||||||||||||| |||:::|||||||||||||
     685 CCACCAGTCTTCAATACTTTAAGATCATTACTACCAAGACAAAGAAAAGT 734

      67 ValAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSe 83
         |||||||||:::|||:::||||||||||||||||||||||||||||||||
     735 GTGGCAATCATAGAAAAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAG 784

      83 rCysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     785 CTGCCGGCTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCA 834

     100 hrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGly 116
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     835 CATTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGT 884

     117 GlnProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrPr 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     885 CAGCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACC 934

     133 oThrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaG 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     935 TACATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTG 984

     150 lyAsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAsp 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     985 GAAATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGAC 1034

     167 ThrAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAl 183
         ||||||||||||||||||||||:::||||||||||||||||||||||||
    1035 ACAGCAGACAGATTTATCAACTCCCAATGTGCAAAATACATGCTAAAAGC 1084

     183 aAsnLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluG 200
         ||||||||||||||||||||||:::|||   |||||||||||||||   |
    1085 CAACCTGATTAAAGAAGCTGAAGACATGAGCTCAAAGTTTACAAGGGGAG 1134

     200 lyThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGln 216
         ||||||||||||||||||||||||||||:::|||||||||||||||:::
    1135 GCACATCAGCGGTAGAGAATTTGAATGAAATTCAGTGCATGTGGTTCCAT 1184

     217 ThrGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLe 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1185 ACAGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACT 1234

     233 uLysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspG 250
         |:::|||||||||||||||||||||||||||||||||||||||||||||||
    1235 TCAGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACC 1284

     250 lnPheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyr 266
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1285 AGTTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATAT 1334
```

```
 267 ValAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPh 283
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1335 GTGGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTT 1384

 283 eLysAla..AlaArgIleAlaIleGluIleTyrLeuLys.LeuHisAspA 299
     ||||:::  |||||||||||||||||||||||||||||| ||||||||||
1385 CAAGGGCAGGCAAGAATTGCTATAGAGATCTATTTGAAGGCTTCATGACA 1434

 299 snProLeuThrAspGluAsn...LysGluHis.GluAla.AspThrAla. 313
     ||||||||||||||||:::   |||||||| |||||| ||||||||||
1435 ACCCCCTTACAGATGAGGAATTAAAGGAACACGGAGGCTGGATACAGCCA 1484

 314 AsnMetSerAspLysGluLeuLysLysLeuArgAsnLysGlnArgArgAl 330
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1485 AACATGTCTGACAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGC 1534

 330 aGlnLysLysAlaGlnIleGluGluGluLysLysAsnAlaGluLysGluL 347
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1535 TCAAAAGAAAGCCCAGATAGAAGAAGAGAAAAAAAAATGCAGAAAAAGAAA 1584

 347 ysProGlnArgAsnProLysLysLysLysAspAspAspAspGluGluIle 363
     ||   |||||||||   ||||||||||||||||||||||||||||||||||
1585 AGCAGCAGAGAAATCAGAAAAAGAAGAAGGATGATGATGATGAGGAGATA 1634

 364 GlyGlyProLysGluGluLeuIleProGluLysLeuAlaLysValGluTh 380
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1635 GGAGGTCCAAAAGAAGAACTTATTCCAGAGAAACTGGCCAAGGTTGAAAC 1684

 380 rProLeuGluGluAlaIleLysPheLeuThrProLeuLysAsnLeuValL 397
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1685 TCCATTGGAAGAAGCTATTAAATTTTTAACACCGTTGAAGAACTTGGTGA 1734

 397 ysAsnLysIleGluThrHisLeuPheAlaPheGluIleTyrPheArgLys 413
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1735 AGAACAAGATAGAGACTCATCTTTTTGCCTTTGAGATTTACTTTAGGAAA 1784

 414 GluLysPheLeuLeuMetLeuGlnSerValLysArgAlaPheAlaIleAs 430

1784 ..................................................... 1784

 430 pSerSerHisProTrpLeuHisGluCysMetIleArgLeuPheHisSerV 447

1784 ..................................................... 1784

 447 alCysGluSerLysAspLeuProGluThrValArgThrValLeuLysGln 463

1784 ..................................................... 1784

 464 GluMetAsnArgLeuPheGlyAlaThrAsnProLysAsnPheAsnGluTh 480

1784 ..................................................... 1784

 480 rPheLeuLysArgAsnSerAspSerLeuProHisArgLeuSerAlaAlaL 497
                                                 |||||||
1785 .....................................GCTGCCA 1791

 497 ysMetValTyrTyrLeuAspSerSerSerGlnLysArgAlaIleGluLeu 513
     ||||||||||||||||||   |||||||||||||||||||||||||||||
1792 AAATGGTATATTACTTAGATCCTTCTAGTCAGAAGCGAGCTATAGAGTTG 1841

 514 AlaThrThrLeuAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMe 530
     ||||||||||||||   ||||||||||||||||||||||||||||||||||
1842 GCAACAACACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTAT 1891

 530 tGluValLeuGluAlaLeuCysAspGlySerLeuArgAspCysLysGluA 547
     ||||||||||||||||||   ||||||||||||   ||||||||||||||
1892 GGAGGTATTGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAG 1941

 547 laAlaGluAlaTyrArgAlaSerCys.HisLysLeuPhe.ProTyrAla. 562
     ||||||||   |||||||||:::||| ||||||||||::: |||||||||
1942 CTGCTGAAATTTATAGAGCAAATTGTCCATAAGCTTTACCCCTTATGCCT 1991
```

```
 563 LeuAlaPhe...MetProProGlyTyrGluGluAspMetLysIle.ThrV 578
     |||||||||      |||||||||||||||||||||||||||||| ||||
1992 TTGGCCTTTCCATGCCCCCCCGGATATGAAGAGGATATGAAGATCCACAG 2041

 578 alAsnGlyAspSerSer.AlaGluThrGluGluLeuAla.AsnGluIle 593
     |||||||||||||||| ||||||:::|||||||||||| |||||||||
2042 TTAATGGAGATAGTTCCTGCAGAAGCTGAAGAACTGGCCCAATGAAATT 2090
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAH77157

seq_documentation_block:

```
ID   AAH77157 standard; cDNA; 1413 BP.
XX
AC   AAH77157;
XX
DT   21-JAN-2002  (first entry)
XX
DE   Human tubedown-1 base pairs 1413-1 antisense cDNA.
XX
KW   Human; tubedown-1; tbdn-1; antisense; cytostatic; osteopathic;
KW   bone tumour; osteosarcoma; Ewings sarcoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179505-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12435.
XX
PR   17-APR-2000; 2000US-197977P.
PR   17-APR-2001; 2001US-0836410.
XX
PA   (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI   Gendron RL,  Paradis H;
XX
DR   WPI; 2002-017618/02.
XX
PT   Nucleic acid molecules antisense to the tubedown-1 gene prevent
PT   overexpression of tubedown-1 protein and are useful to treat
PT   osteosarcoma and Ewing's Sarcoma family of tumours  -
XX
PS   Claim 6; Page 38-39; 56pp; English.
XX
CC   The sequence represents tubedown-1 (tbdn-1) bases 1413-1 antisense cDNA.
CC   The invention relates to a novel isolated nucleic acid of the tubedown-1
CC   gene; and antisense nucleic acids to tbdn-1. The polynucleotides and
CC   protein of the invention have cytostatic and osteopathic activity. The
CC   polynucleotides of the invention may be used in antisense-therapy/gene
CC   therapy. They are useful in the treatment of bone tumours, especially
CC   osteosarcoma and Ewings sarcoma family of tumours. The compounds of the
CC   invention may also be useful for the prevention of metastases from these
CC   types of tumours, either alone or in combination with radiotherapy and/or
CC   chemotherapeutic agents.
XX
SQ   Sequence 1413 BP; 359 A; 314 C; 244 G; 496 T; 0 other;
```

alignment_scores:

```
           Quality: 1768.00              Length:    335
             Ratio:  5.278                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-836-410A-1 x AAH77157/rev ..

Align seg 1/1 to reverse of: AAH77157 from: 1 to: 1413

```
   1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1006 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 957
```

```
 17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
    |||||||||||||||||||||||||||||||||||||||||||||||||||
956 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTTATCTGGAGAGAAGT 907

 34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
    |||||||||||||||||||||||||||||||||||||||||||||||||||
906 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 857

 51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
    |||||||||||||||||||||||||||||||||||||||||||||||||||
856 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 807

 67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
    |||||||||||||||||||||||||||||||||||||||||||||||||||
806 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 757

 84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
    |||||||||||||||||||||||||||||||||||||||||||||||||||
756 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
    |||||||||||||||||||||||||||||||||||||||||||||||||||
706 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 657

117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
    |||||||||||||||||||||||||||||||||||||||||||||||||||
656 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 607

134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
    |||||||||||||||||||||||||||||||||||||||||||||||||||
606 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 557

151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
    |||||||||||||||||||||||||||||||||||||||||||||||||||
556 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 507

167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
    |||||||||||||||||||||||||||||||||||||||||||||||||||
506 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 457

184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
456 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 407

201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
    |||||||||||||||||||||||||||||||||||||||||||||||||||
406 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 357

217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
    |||||||||||||||||||||||||||||||||||||||||||||||||||
356 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 307

234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
306 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 257

251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
    |||||||||||||||||||||||||||||||||||||||||||||||||||
256 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 207

267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
    |||||||||||||||||||||||||||||||||||||||||||||||||||
206 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 157

284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
156 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 107

301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
    |||||||||||||||||||||||||||||||||||||||||||||||||||
106 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 57

317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
```

```
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 56 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 7

334 laGln 335
    |||||
  6 CCCAG 2
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH16408

seq_documentation_block:

ID AAH16408 standard; cDNA; 1802 BP.
XX
AC AAH16408;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15380.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15380; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1802 BP; 644 A; 298 C; 395 G; 465 T; 0 other;

```
alignment_scores:
            Quality: 1747.00              Length:    336
              Ratio:  5.215                Gaps:      0
 Percent Similarity:  99.702   Percent Identity: 98.512

alignment_block:
 US-09-836-410A-1 x AAH16408   ..

 Align seg 1/1  to: AAH16408  from: 1  to: 1802

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||   ||||||||||||||
     794 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTAGACTAAATATCCCAG 843

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     844 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 893

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
     894 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 943

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
     944 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 993

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         |||||||:::|||||||||||||||||||||||||||||||||||||||||
     994 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1043

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1044 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1093

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1094 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1143

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1144 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1193

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1194 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1243

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1244 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1293

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1294 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1343

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1344 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1393

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1394 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1443

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1444 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1493

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1494 AGAAATGTCACGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAG 1543

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||

    1544 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1593

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1594 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1643

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1644 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCC 1693

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1694 CTTACAGATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGA 1743

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1744 CAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1793

     334 laGlnIle 336
         ||||||||
    1794 CCCAGATA 1801

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH16424

seq_documentation_block:
ID   AAH16424 standard; cDNA; 1985 BP.
XX
AC   AAH16424;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:15407.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 15407; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
```

```
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1985 BP; 652 A; 349 C; 492 G; 492 T; 0 other;


alignment_scores:
            Quality: 1521.50              Length:    298
              Ratio:  5.175                Gaps:      1
 Percent Similarity:  98.658   Percent Identity: 97.315

alignment_block:
 US-09-836-410A-1 x AAH16424    ..

 Align seg 1/1  to: AAH16424  from: 1  to: 1985

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1096 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTGGACTAAATATCCCAG 1145

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||:::||||||||||||||||||||||||||||||
    1146 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 1195

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||:::||||||||||||||||||||||||||||||
    1196 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 1245

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||:::||||||||||||||
    1246 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 1295

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         |||||||:::||||||||||||||||||||||||||||||||||||||||
    1296 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1345

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1346 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1395

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1396 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1445

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1446 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1495

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1496 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1545

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1546 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1595

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1596 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1645

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1646 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1695

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1696 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1745

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1746 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1795

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||          ||||||||||||||||
    1796 AGAAATGTCATGAGATTGAGAGAAA.........ATCACTGATGACCAG 1835

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1836 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1885

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1886 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1935

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAsp 298
         |||||||||||||||||||||||||||||||||||||||||||||
    1936 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGAC 1979

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH81664

seq_documentation_block:
ID   AAH81664 standard; DNA; 710 BP.
XX
AC   AAH81664;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human differential transcription-associated cDNA SEQ ID 173.
XX
KW   Differential transcription; human; rat; tumour cell; cytostatic;
KW   Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157058-A2.
XX
PD   09-AUG-2001.
XX
PF   31-JAN-2001; 2001WO-EP01003.
XX
PR   31-JAN-2000; 2000DE-1004102[redacted]
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
PI   Grips M, Hellriegel M, Schmitz A, Sers C;
XX
DR   WPI; 2001-483415/52.
XX
PT   Nucleic acids differentially expressed between tumor and normal cells,
PT   useful for diagnosis or therapy of tumors and for screening active
PT   agents -
XX
PS   Disclosure; Page 376; 579pp; German.
XX
CC   This invention describes a nucleic acid (I) with differential expression
CC   between tumour and normal cells and which has cytostatic activity. (I)
CC   work as modulators of Ras activity by inducing expression of tumour
CC   suppressor genes. (I), and polypeptides encoded by them, are useful as
CC   targets for diagnosis or therapy and in screening to determine the
CC   effects of an active compound (potential pharmaceutical) on a cell line,
CC   particularly for diagnosis and treatment of tumors, especially by
CC   modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC   methods) or by modulating the amount and/or location of (I)-encoded
CC   polypeptides (by administration of the polypeptide or its activator,
CC   antibody (optionally as a conjugate) or inhibitor). The method allows
```

```
CC   identification of many Class II tumour suppressor genes (i.e. genes that
CC   are not primary targets for tumour-initiating mutations).
CC   AAH81492-AAH82376 represent the human and rat derived nucleic acid
CC   fragments described in the method of the invention.
XX
SQ   Sequence 710 BP; 170 A; 146 C; 121 G; 267 T; 6 other;


alignment_scores:
           Quality: 1176.00              Length:    237
             Ratio:  5.091                Gaps:      0
 Percent Similarity:  97.468    Percent Identity: 96.203

alignment_block:
 US-09-836-410A-1 x AAH81664/rev   ..

 Align seg 1/1  to reverse of: AAH81664  from: 1  to: 710

     103 TrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSe 119
         |||||||:::|||||||||||||:::|||||||||||||||||||||||||
     710 TGGGTCAAGTACTACTTGGCACACCATTATGACAAAATTGGTCAGCCATC 661

     119 rIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuI 136
         |   |||   |||||||||||||||||||||   ||||||   |||||||
     660 TANTGCTNTGGAGTACATAAATACTGCTATTGANAGTACACT.ACATTAA 612

     136 leGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIle 152
         ||:::   |||||||||   |||||||||||||||||||||||||||||||
     611 TAGACNTCTTTCTCGTGANAGCTAAAATCTATAAGCATGCTGGAAATATT 562

     153 LysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAs 169
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     561 AAAGAAGCTGCAAGGTGGATGGATGAGGCNCAGGCCTTGGACACAGCAGA 512

     169 pArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuI 186
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     511 CAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGA 462

     186 leLysGluAlaGluGluMetCysSerLysPheThrArgGluGlyThrSer 202
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     461 TTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGAACATCA 412

     203 AlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnThrGluCy 219
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     411 GCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAACAGAATG 362

     219 sAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuLysLysC 236
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     361 TGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTAAGAAAT 312

     236 ysHisGluIleGluArgHisPheIleGluIleThrAspAspGlnPheAsp 252
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     311 GTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAGTTTGAC 262

     253 PheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLe 269
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGTGGACTT 212

     269 uLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaA 286
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     211 ATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCAAGGCAG 162

     286 laArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThr 302
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     161 CAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCCCTTACA 112

     303 AspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGl 319
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     111 GATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGACAAAGA 62

     319 uLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnI 336
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      61 GCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAGCCCAGA 12

     336 leGluGluGlu 339
         ||||||||||||
      11 TAGAAGAAGAG 1

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ15705

seq_documentation_block:
ID   AAZ15705 standard; cDNA; 781 BP.
XX
AC   AAZ15705;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:3174.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1524-1525; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 781 BP; 263 A; 140 C; 156 G; 205 T; 17 other;
```

```
alignment_scores:
           Quality: 1104.00              Length:    232
             Ratio:  4.973                Gaps:      4
 Percent Similarity:  95.690   Percent Identity: 95.259

alignment_block:
 US-09-836-410A-1 x AAZ15705   ..

 Align seq 1/1  to: AAZ15705  from: 1  to: 781

      62 ArgAspLysGlu.LysValAlaIleValGluGluLeuValValGlyTyrG 78
         |||||||||||| |||||||||||:::|||||||||||||||||||||||
      87 AGAGACAAAGAAAAAGGTGGCAATCATAGAAGAGTTAGTAGTAGGTTATG 136

      78 luThrSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLys 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     137 AAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATGATGGAAAG 186

      95 GluGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHi 111
         |||||||||||||||||||||||||||||||      ||||||||||||||
     187 GAGGAACCACCAACCACATTACTTTGGGTCCNNTNCTACTTGGCACAACA 236

     111 sTyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrA 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     237 TTATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTG 286

     128 laIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLys 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     287 CTATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAA 336

     145 IleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGl 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     337 ATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGA 386

     161 uAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaL 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     387 GGCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAA 436

     178 ysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSer 194
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     437 AATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCA 486

     195 LysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGl 211
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     487 AAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCA 536

     211 nCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnL 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     537 GTGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATA 586

     228 ysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIle 244
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     587 AATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACATTTTATA 636

     245 .GluIleThrAspAspGlnPheAspPheHisThrTyrCysMet.ArgLys 260
          |||||||||||||||||||||||||||||||||||   ||| ||||||
     637 GGAAATCACTGATGACCAGTTTGACTTTCATACATACTGGATGAAGGAAG 686

     261 IleThrLeuArgSerTyrValAspLeuLeuLysLeu.GluAspValLeuA 277
         |||||||||||||||||||||||||||    |||||| ||||||||||||
     687 ATTACCCTTAGATCATATGTGGACTTATTNAAACTATGAAGATGTACTTT 736

     277 rgGlnHisProPheTyrPheLysAlaAlaArgIleAla 289
         ||||||   |||||||||||||||||||||||||||
     737 NACAGCATNCATTTTACTTCAAGGCAGCAAGAATTGCT 774

 seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX99053

 seq_documentation_block:
 ID   AAX99053 standard; cDNA; 781 BP.
 XX
```

```
AC   AAX99053;
XX
DT   24-SEP-1999  (first entry)
XX
DE   Human validated cancer cell derived cDNA #375.
XX
KW   Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW   integral membrane protein; aspartyl protease; GATA family; wnt family;
KW   transcription factor; G-protein alpha subunit; protein phosphatase;
KW   phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW   protein kinase; tyrosine phosphatase; developmental signalling protein;
KW   WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW   detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW   Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW   leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW   prostate; ss.
XX
OS   Homo sapiens.
XX
PN   WO9933982-A2.
XX
PD   08-JUL-1999.
XX
PF   22-DEC-1998;   98WO-US27610.
XX
PR   21-DEC-1998;   98US-0217471.
PR   23-DEC-1997;   97US-0068755.
PR   03-APR-1998;   98US-0080664.
PR   21-OCT-1998;   98US-0105234.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones LW,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-430243/36.
XX
PT   New isolated human polynucleotides
XX
PS   Claim 1; Page 564; 591pp; English.
XX
CC   This invention describes novel isolated human polynucleotides obtained
CC   by screening for differential expression in colon cancer, breast cancer
CC   and lung cancer cell lines. The polynucleotides of the invention are
CC   represented in AAX98275-X99118 and encode polypeptides of protein
CC   families selected from 4 transmembrane segments integral membrane
CC   proteins, 7 transmembrane receptors, ATPases associated with various
CC   cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC   transcription factors, G-protein alpha subunit, phorbolesters or
CC   diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC   protein tyrosine phosphatase, trypsin, wnt family of developmental
CC   signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC   encoded polypeptides also have a functional domain selected from Ank
CC   repeat, basic region plus leucine zipper transcription factors,
CC   bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC   (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC   domain. The polynucleotides encode polypeptides with similarity to known
CC   protein families and are predicted to have similar properties. The novel
CC   polynucleotides can be used to develop products for use as therapeutic
CC   agents and in forensics, genetic analysis, mapping and diagnostic
CC   applications. In particular, the product can be used for the detection
CC   and management of cancers. They can be used for treating e.g. cervical
CC   cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC   retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC   myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC   myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC   hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC   epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
```

```
CC   mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC   prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC   the skin.
XX
SQ   Sequence 781 BP; 263 A; 140 C; 156 G; 205 T; 17 other;


alignment_scores:
           Quality: 1104.00              Length:    232
             Ratio:  4.973                Gaps:      4
 Percent Similarity:  95.690   Percent Identity: 95.259

alignment_block:
 US-09-836-410A-1 x AAX99053    ..

 Align seg 1/1  to: AAX99053  from: 1  to: 781

      62 ArgAspLysGlu.LysValAlaIleValGluGluLeuValValGlyTyrG 78
         ||||||||||| |||||||||||:::|||||||||||||||||||||||
      87 AGAGACAAAGAAAAAGGTGGCAATCATAGAAGAGTTAGTAGTAGGTTATG 136

      78 luThrSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLys 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     137 AAACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATGATGGAAAG 186

      95 GluGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHi 111
         |||||||||||||||||||||||||||||      ||||||||||||||
     187 GAGGAACCACCAACCACATTACTTTGGGTCCNNTNCTACTTGGCACAACA 236

     111 sTyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrA 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     237 TTATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTG 286

     128 laIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLys 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     287 CTATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAA 336

     145 IleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGl 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     337 ATCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGA 386

     161 uAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaL 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     387 GGCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAA 436

     178 ysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSer 194
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     437 AATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCA 486

     195 LysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGl 211
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     487 AAGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCA 536

     211 nCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnL 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     537 GTGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATA 586

     228 ysPheGlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIle 244
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     587 AATTTGGTGAAGCACTTAAGAAATGTCATGAGATTGAGAGACATTTTATA 636

     245 .GluIleThrAspAspGlnPheAspPheHisThrTyrCysMet.ArgLys 260
         |||||||||||||||||||||||||||||||||||   ||| ||||||
     637 GGAAATCACTGATGACCAGTTTGACTTTCATACATACTGGATGAAGGAAG 686

     261 IleThrLeuArgSerTyrValAspLeuLeuLysLeu.GluAspValLeuA 277
         |||||||||||||||||||||||||||||    |||||| ||||||||||||
     687 ATTACCCTTAGATCATATGTGGACTTATTNAAACTATGAAGATGTACTTT 736

     277 rgGlnHisProPheTyrPheLysAlaAlaArgIleAla 289
           ||||||   |||||||||||||||||||||||||||
     737 NACAGCATNCATTTTACTTCAAGGCAGCAAGAATTGCT 774
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ15983

seq_documentation_block:
ID   AAZ15983 standard; cDNA; 764 BP.
XX
AC   AAZ15983;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:3452.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1650-1651; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 764 BP; 253 A; 138 C; 148 G; 204 T; 21 other;


alignment_scores:
           Quality:  973.00              Length:    200
             Ratio:  5.015                Gaps:      3
```

```
 Percent Similarity:  97.000   Percent Identity: 95.500

alignment_block:
 US-09-836-410A-1 x AAZ15983   ..

 Align seq 1/1  to: AAZ15983  from: 1  to: 764

      62 ArgAspLysGluLysValAlaIleValGluGluLeuValValGlyTyrGl 78
         ||||||||||||||||||||||||:::||||||:::   ||||||||||
      65 AGAGACAAAGAAAAGGTGGCAATCATAGAAGAGTTTNTAGTAGGTTATGA 114

      78 uThrSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLysG 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 AACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATGATGGAAAGG 164

      95 luGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHis 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 AGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACAT 214

     112 TyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrAl 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 TATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGC 264

     128 aIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLysI 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 TATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAA 314

     145 leTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGlu 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 TCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAG 364

     162 AlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaLy 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 GCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAA 414

     178 sTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerL 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 ATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAA 464

     195 ysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGln 211
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 AGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAG 514

     212 CysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsn.L 228
         ||||||||||||||||||||||||||||||||||||||||||||||||| |
     515 TGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATTA 564

     228 ysPheGlyGluAlaLeuLysLysCysHis.GluIleGluArgHisPheIl 244
         |||||||||||||||||||||||||||||| ||||||||||||   :::
     565 AATTTGGTGAAGCACTTAAGAAATGTCATTGAGATTGAGAGACTTTTATA 614

     244 eGluIleThrAspAsp.GlnPheAspPheHisThrTyrCysMet 258
         ||||||||||||||| |||||||||||||||||||||||||||
     615 GGAAATCACTGATGACCCAGTTTGACTTTCATACATACTGTATG 658

 seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX98777

 seq_documentation_block:
 ID   AAX98777 standard; cDNA; 764 BP.
 XX
 AC   AAX98777;
 XX
 DT   24-SEP-1999  (first entry)
 XX
 DE   Human validated cancer cell derived cDNA #99.
 XX
 KW   Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
 KW   integral membrane protein; aspartyl protease; GATA family; wnt family;
 KW   transcription factor; G-protein alpha subunit; protein phosphatase;
 KW   phorbolester binding protein; diacylglycerol binding protein; trypsin;
 KW   protein kinase; tyrosine phosphatase; developmental signalling protein;
 KW   WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
 KW   detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
 KW   Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
 KW   leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
 KW   prostate; ss.
 XX
 OS   Homo sapiens.
 XX
 PN   WO9933982-A2.
 XX
 PD   08-JUL-1999.
 XX
 PF   22-DEC-1998;   98WO-US27610.
 XX
 PR   21-DEC-1998;   98US-0217471.
 PR   23-DEC-1997;   97US-0068755.
 PR   03-APR-1998;   98US-0080664.
 PR   21-OCT-1998;   98US-0105234.
 PR   27-OCT-1998;   98US-0105877.
 XX
 PA   (CHIR ) CHIRON CORP.
 PA   (HYSE-) HYSEQ INC.
 XX
 PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
 PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
 PI   Jones LW,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
 PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
 PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
 XX
 DR   WPI; 1999-430243/36.
 XX
 PT   New isolated human polynucleotides
 XX
 PS   Claim 1; Page 453; 591pp; English.
 XX
 CC   This invention describes novel isolated human polynucleotides obtained
 CC   by screening for differential expression in colon cancer, breast cancer
 CC   and lung cancer cell lines. The polynucleotides of the invention are
 CC   represented in AAX98275-X99118 and encode polypeptides of protein
 CC   families selected from 4 transmembrane segments integral membrane
 CC   proteins, 7 transmembrane receptors, ATPases associated with various
 CC   cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
 CC   transcription factors, G-protein alpha subunit, phorbolesters or
 CC   diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
 CC   protein tyrosine phosphatase, trypsin, wnt family of developmental
 CC   signalling proteins and WW/rsp5/WWP domain containing proteins. The
 CC   encoded polypeptides also have a functional domain selected from Ank
 CC   repeat, basic region plus leucine zipper transcription factors,
 CC   bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
 CC   (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
 CC   domain. The polynucleotides encode polypeptides with similarity to known
 CC   protein families and are predicted to have similar properties. The novel
 CC   polynucleotides can be used to develop products for use as therapeutic
 CC   agents and in forensics, genetic analysis, mapping and diagnostic
 CC   applications. In particular, the product can be used for the detection
 CC   and management of cancers. They can be used for treating e.g. cervical
 CC   cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
 CC   retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
 CC   myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
 CC   myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
 CC   hereditary ectodermal dysplasia, congenital alveolar dysplasia,
 CC   epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
 CC   mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
 CC   prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
 CC   the skin.
 XX
 SQ   Sequence 764 BP; 253 A; 138 C; 148 G; 204 T; 21 other;


 alignment_scores:
           Quality:  973.00            Length:    200
             Ratio:  5.015               Gaps:      3
  Percent Similarity:  97.000   Percent Identity: 95.500

 alignment_block:
```

```
US-09-836-410A-1 x AAX98777   ..

Align seg 1/1  to: AAX98777  from: 1  to: 764

     62 ArgAspLysGluLysValAlaIleValGluGluLeuValValGlyTyrGl 78
        ||||||||||||||||||||||||||:::|||||:::   ||||||||||
     65 AGAGACAAAGAAAAGGTGGCAATCATAGAAGAGTTTNTAGTAGGTTATGA 114

     78 uThrSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLysG 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    115 AACCTCTCTAAAAAGCTGCCGGTTATTTAACCCCAATGATGATGGAAAGG 164

     95 luGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHis 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    165 AGGAACCACCAACCACATTACTTTGGGTCCAGTACTACTTGGCACAACAT 214

    112 TyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrAl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    215 TATGACAAAATTGGTCAGCCATCTATTGCTTTGGAGTACATAAATACTGC 264

    128 aIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLysI 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    265 TATTGAAAGTACACCTACATTAATAGAACTCTTTCTCGTGAAAGCTAAAA 314

    145 leTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGlu 161
        ||||||||||||||||||||||||||||||||||||||||||||||||
    315 TCTATAAGCATGCTGGAAATATTAAAGAAGCTGCAAGGTGGATGGATGAG 364

    162 AlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaLy 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 GCCCAGGCCTTGGACACAGCAGACAGATTTATCAACTCCAAATGTGCAAA 414

    178 sTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerL 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    415 ATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAAATGTGCTCAA 464

    195 ysPheThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGln 211
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 AGTTTACAAGGGAAGGAACATCAGCGGTAGAGAATTTGAATGAAATGCAG 514

    212 CysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsn.L 228
        |||||||||||||||||||||||||||||||||||||||||||||||| |
    515 TGCATGTGGTTCCAAACAGAATGTGCCCAGGCTTATAAAGCAATGAATTA 564

    228 ysPheGlyGluAlaLeuLysLysCysHis.GluIleGluArgHisPheIl 244
        |||||||||||||||||||||||||||||| ||||||||||||   :::
    565 AATTTGGTGAAGCACTTAAGAAATGTCATTGAGATTGAGAGACTTTTATA 614

    244 eGluIleThrAspAsp.GlnPheAspPheHisThrTyrCysMet 258
        |||||||||||||||| |||||||||||||||||||||||||||
    615 GGAAATCACTGATGACCCAGTTTGACTTTCATACATACTGTATG 658

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS37350

seq_documentation_block:
ID   AAS37350 standard; cDNA; 402 BP.
XX
AC   AAS37350;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #408.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A,  Lamson G;
PI   Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones WL,  Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 698-699; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 402 BP; 142 A; 76 C; 85 G; 99 T; 0 other;


alignment_scores:
            Quality:  680.00              Length:    130
              Ratio:  5.231                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-836-410A-1 x AAS37350   ..

 Align seg 1/1  to: AAS37350  from: 1  to: 402

      71 GluGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPh 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      12 GAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCTGCCGGTTATT 61

      87 eAsnProAsnAspAspGlyLysGluGluProProThrThrLeuLeuTrpV 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      62 TAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACATTACTTTGGG 111

     104 alGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnProSerIle 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     112 TCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCAGCCATCTATT 161

     121 AlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrLeuIleGl 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     162 GCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTACATTAATAGA 211

     137 uLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsnIleLysG 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     212 ACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGAAATATTAAAG 261

     154 luAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAlaAspArg 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     262 AAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACACAGCAGACAGA 311

     171 PheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaAsnLeuIleLy 187
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     312 TTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCAACCTGATTAA 361

     187 sGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         |||||||||||||||||||||||||||||||||||||
     362 AGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 401

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH12222
```

```
seq_documentation_block:
ID   AAH12222 standard; cDNA; 488 BP.
XX
AC   AAH12222;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9057.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 9057; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 488 BP; 121 A; 98 C; 84 G; 167 T; 18 other;


alignment_scores:
            Quality:  605.50            Length:    157
              Ratio:  4.388               Gaps:      1
 Percent Similarity:  87.898   Percent Identity: 82.166

alignment_block:
 US-09-836-410A-1 x AAH12222/rev   ..

Align seg 1/1  to reverse of: AAH12222  from: 1  to: 488

    142 LysAlaLysIleTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTr 158
        :::|||||||||||||||   |||         |||||||||:::::
    466 GAAGCTAAAATCTATAAGCA.GCTGNAANTTTNAAAGAAGCCTNCAAGGG 418

    158 pMetAspGluAlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerL 175
        ||||||   |||||||||          :::::::::|||||||||||||
    417 GATGGATTAGGCCCAGGCCNTTGGACCCAGCAGCCAGTTTATCAACTCCA 368

    175 ysCysAlaLysTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGlu 191
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    367 AATGTGCAAAATACATGCTAAAAGCCAACCTGATTAAAGAAGCTGAAGAA 318

    192 MetCysSerLysPheThrArgGluGlyThrSerAlaValGluAsnLeuAs 208
        ||||||||||||||:::|||||||||||||||||||||   ||||||||
    317 ATGTGCTCAAAGTTTNCAAGGGAAGGAACATCAGCGGTANAGAATTTGAA 268

    208 nGluMetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysA 225
        ||||||||||   |||   ||||||||||||   ||||||||||||||||
    267 TGAAATGCAGNGCATGNGGTTCCAAACAGAATGNGCCCAGGCTTATAAAG 218

    225 laMetAsnLysPheGlyGluAlaLeuLysLysCysHisGluIleGluArg 241
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 CAATGAATAAATTTGGTGAAGCNCTTAAGAAATGTCATGAGATTGAGAGA 168

    242 HisPheIleGluIleThrAspAspGlnPheAspPheHisThrTyrCysMe 258
                |||:::|||||||||||||||||||||||||||||||||
    167 AA.........ATCNCTGATGACCAGTTTGACTTTCATACATACTGTAT 128

    258 tArgLysIleThrLeuArgSerTyrValAspLeuLeuLysLeuGluAspV 275
        ||||||||||||||||||||||||   ||||||||||||||||||||
    127 GAGGAAGATTACCCTTAGATCATATGNGGACTTATTAAAACTAGAAGATG 78

    275 alLeuArgGlnHisProPheTyrPheLysAlaAlaArgIleAlaIleGlu 291
         ||||||||||||||||||||||||||||||||||||||||||||||||
     77 NACTTCGACAGCATCCNTTTTACTTCAAGGCAGCAAGAATTGCTATANAG 28

    292 IleTyrLeuLysLeuHisAsp 298
        |||||||||||||||||||
     27 ATCTATTTGAAGCTTCATGAC 7

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:AAN90541

seq_documentation_block:
ID   AAN90541 standard; DNA; 2703 BP.
XX
AC   AAN90541;
XX
DT   28-NOV-1989  (first entry)
XX
DE   DNA encoding N-alpha-acetyl transferase.
XX
KW   N-alpha-acetyl transferase; herbicide resistance;
KW   protein N-acetylation.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    272..335
FT                   /*tag=  a
FT   misc_feature    338..392
FT                   /*tag=  b
FT   misc_feature    479..515
FT                   /*tag=  c
FT   misc_feature    542..566
FT                   /*tag=  d
FT   misc_feature    971..989
FT                   /*tag=  e
FT   misc_feature    1007..1049
FT                   /*tag=  f
FT   misc_feature    1061..1085
FT                   /*tag=  g
FT   misc_feature    1088..1130
```

```
FT                   /*tag=  h
FT   misc_feature    1481..1505
FT                   /*tag=  i
FT   misc_feature    1829..1856
FT                   /*tag=  j
FT   misc_feature    1862..1885
FT                   /*tag=  k
FT   misc_feature    1909..1934
FT                   /*tag=  l
FT   misc_feature    2072..2117
FT                   /*tag=  m
FT   misc_feature    2123..2183
FT                   /*tag=  n
XX
PN   WO8907138-A.
XX
PD   10-AUG-1989.
XX
PF   07-FEB-1989;    89WO-US00471.
XX
PR   08-FEB-1988;    88US-0153361.
PR   14-DEC-1988;    88US-0284344.
XX
PA   (GEHO ) THE GENERAL HOSPITAL CORPORATION.
XX
PI   Smith JA,  Lee FJS;
XX
DR   WPI; 1989-249008/34.
DR   P-PSDB; AAP91070.
XX
PT   New pure N-alpha-acetyl transferase and DNA encoding it - catalysing
PT   acetylation of proteins and peptides, eg to stabilise pharmaceuticals
PT   or induce herbicide resistance in plants.
XX
PS   Claim 8; Page 50; fig  12b-e; 72pp; English.
XX
CC   DNA encodes N-alpha-actyl transferase, used for catalysing N-acetylation
CC   of peptides/proteins, eg to stabilise pharmaceuticals or to induce
CC   herbicide resistance in plants. Features a - n are fragments resulting
CC   from exonuclease III deletion. See also AAP91070.
XX
SQ   Sequence 2703 BP; 943 A; 489 C; 530 G; 741 T; 0 other;


alignment_scores:
            Quality:  550.50              Length:    595
              Ratio:  1.605                Gaps:     17
 Percent Similarity:  57.647   Percent Identity: 27.059

alignment_block:
US-09-836-410A-1 x AAN90541   ..

 Align seg 1/1  to: AAN90541  from: 1  to: 2703

      15 TyrProArgGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGl 31
         |||||||||          |||:::  :::|||||:::|||||||:::
     913 TATCCAAGATGCGAACCACCCAAATTTATTCCATTAACTTTCCTTCAAGA 962

      31 y...GluLysPheLysGluCysLeuAspArgPheLeuArgMetAsnPheS 47
            |||:::::::::::::   |||   :::::::::      :::::::
     963 CAAAGAAGAGCTCAGCAAAAAATTGAGAGAATATGTTTTGCCTCAATTGG 1012

      47 erLysGlyCysProProValPheAsnThrLeuArgSerLeuTyr...Arg 62
         :::::|||   |||   :::|||:::::::::::   ||||||   |||
    1013 AGCGCGGTGTTCCAGCAACTTTTTCCAACGTGAAACCCCTTTACCAAAGA 1062

      63 AspLysGluLysValAla...IleValGluGluLeuValValGlyTyrGl 78
            |||:::||||||:::   :::::::|||:::::::|||:::   |||
    1063 AGGAAGTCCAAGGTTTCACCACTATTGGAGAAAATTGTCCTTGATTATTT 1112

      78 uThrSerLeuLysSerCysArgLeuPheAsnProAsnAspAspGlyLysG 95
         :::::::|||                 :::|||:::::|||
    1113 GTCCGGATTA.................GATCCTACGCAGGAT....... 1137

      95 luGluProProThrThrLeuLeuTrpValGlnTyrTyrLeuAlaGlnHis 111
            |||      ::::::|||:::::::|||||||||:::||||||
    1138 ........CCAATTCCTTTTATTTGGACCAATTATTACTTGTCTCAACAT 1179

     112 TyrAspLysIleGlyGlnProSerIleAlaLeuGluTyrIleAsnThrAl 128
         :::      :::   :::         |||   |||||||||:::::||
    1180 TTCCTTTTCCTTAAGGATTTTCCGAAAGCCCAAGAATATATTGATGCTGC 1229

     128 aIleGluSerThrProThrLeuIleGluLeuPheLeuValLysAlaLysI 145
         |::::::   ||||||||||||:::|||:::::::::::::|||||:::|
    1230 CCTTGACCACACCCCAACTTTAGTTGAGTTTTACATCCTCAAGGCACGTA 1279

     145 leTyrLysHisAlaGlyAsnIleLysGluAlaAlaArgTrpMetAspGlu 161
         ||   ||||||   |||   :::      ||||||      :::::|||
    1280 TCCTGAAGCACTTAGGCCTAATGGACACAGCGGCTGGAATTTTGGAGGAA 1329

     162 AlaGlnAlaLeuAspThrAlaAspArgPheIleAsnSerLysCysAlaLy 178
         ::::::   ||||||      ||||||||||||||   |||   :::||
    1330 GGTAGGCAACTTGATTTGCAGGATAGATTTATCAACTGTAAAACGGTTAA 1379

     178 sTyrMetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerL 195
         ||||:::|||:::||||||   |||   ::|||   |||:::::::|||
    1380 GTACTTTTAAGGGCTAACAATATCGACAAGGCGGTGGAAGTCGCGTCCC 1429

     195 ysPheThrArg......GluGlyThrSerAlaValGluAsnLeuAsnGlu 209
         . ||||||:::      :::::::::::::::::::::::|||:::
    1430 TTTTCACCAAAAACGATGATTCTGTTAATGGTATTAAGGACTTACACCTT 1479

     210 MetGlnCysMetTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMe 226
         :::::::::   ||||||   :::|||   |||:::||||||      ::
    1480 GTGGAAGCTTCTTGGTTCATCGTAGAACAGGCAGAAGCCTATTATAGACT 1529

     226 t..................................................... 226
         :
    1530 ATACCTGGATAGAAAAAAGAAATTAGACGATTTAGCATCGCTAAAAAAAG 1579

     226 ..................................................... 226
         .
    1580 AGGTTGAAAGTGATAAAAGCGAACAAATTGCGAATGATATCAAAGAAAAC 1629

     227 ...........AsnLysPhe...GlyGluAlaLeuLysLysCysHisGl 238
                 :::|||:::   |||   ||||||||||:::   :::
    1630 CAATGGCTTGTTCGCAAATATAAAGGTTTGGCGCTGAAAAGATTCAACGC 1679

     238 uIleGluArgHisPheIleGluIleThrAspAspGlnPheAspPheHisT 255
         '|||   :::   :::   ::::::   ||||||||||:::||||||||:
    1680 TATTCCAAAGTTTTATAAACAATTCGAAGATGACCAGTTGGATTTCCATT 1729

     255 hrTyrCysMetArgLysIleThrLeuArgSerTyrValAspLeuLeuLys 271
         ::|||||||||||||||   |||   |||:::|||:::::::::|||:::
    1730 CATACTGTATGAGAAAAGGTACGCCAAGAGCCTATCTGGAGATGTTAGAA 1779

     272 LeuGluAspValLeuArgGlnHisProPheTyrPheLysAlaAlaArgIl 288
         .       :::|||         |||:::|||   :::|||   :::
    1780 TGGGGAAAGGCACTTTATACCAAACCCATGTATGTTCGCGCAATGAAGGA 1829

     288 eAlaIleGluIleTyrLeuLysLeuHisAspAsnProLeuThrAspGluA 305
         '|||   :::::::|||:::::::::|||||:::   |||
    1830 AGCATCAAAGCTTTACTTTCAAATGCATGATGATCGCTTA.......... 1869

     305 snLysGluHisGluAlaAspThrAlaAsnMetSerAspLysGluLeuLys 321
            :::::::::::::|||:::   :::   ::::::   |||:::::::
    1870 .....AAAAGAAAGTCCGATTCTTTAGATGAAAATTCAGATGAAATCCAA 1914

     322 LysLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGl 338
         :::      :::::::::          :::||||||||||:::   |||
    1915 AATAATGGCCAAAATAGTAGCAGCCAAAAGAAAAAAGCTAAGAAGGAAGC 1964

     338 uGluLysLysAsnAlaGluLysGluLysProGlnArgAsnProLysLysL 355
            |||   :::|||||||         ::::::
    1965 AGCCGCTATGAACAAACGGAAAGAAACTGAAGCCAAGAGTGTTGCTGCTT 2014
```

```
 355 ysLysAspAspAspAspGluGluIleGlyGlyProLysGluGluLeuIle 371
       :::|||:::|||:::::::::   |||
2015 ATCCGAGTGATCAAGATAACGATGTATTCGGC..................... 2046

 372 ProGluLysLeuAlaLysValGluThrProLeuGluGlu...AlaIleLy 387
     |||||||||   :::::::::|||||||:::|||:::   |||   ::
2047 ...GAAAAGTTGATTGAAACCTCCACTCCAATGGAGGACTTCGCTACCGA 2093

 387 sPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluThrHisL 404
     :|||   :::      :::      |||:::::::      :::   ::::
2094 ATTTTATAATAACTACTCCATGCAAGTCAGAGAAGACGAAAGGGATTATA 2143

 404 euPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeuMetLeu 420
     :::::   |||||||:::   :::|||      |||:::   |||   :::
2144 TTTTGGACTTTGAATTTAACTACAGAATTGGAAAGTTAGCTTTGTGCTTT 2193

 421 GlnSerVal...LysArgAlaPheAlaIleAspSerSerHisProTrpLe 436
     |||:::   |||   |||      :::   ::::::      ::
2194 GCTTCTCTAAACAAATTCGCTAAGAGATTTGGCACCACGAGCGGTTTATT 2243

 436 uHisGluCysMetIleArgLeuPheHisSerValCysGluSerLysAspL 453
     :   :::      |||   |||:::|||::::::   ::::::      :
2244 TGGTAGTATGGCCATTGTTTTGTTACATGCCACAAGAAACGACACCCCCT 2293

 453 euProGluThrValArgThrValLeuLysGlnGluMetAsnArgLeuPhe 469
     ::         ::::::   |||:::   ::::::::::::::   :::
2294 TTGATCCAATTTTGAAGAAAGTAGTCACGAAGAGCCTTGAAAAAGAGTAT 2343

 470 GlyAlaThrAsnProLysAsn..................................... 476
     :::   :::   |||   |||
2344 TCTGAAAATTTCCCATTAAACGAAATATCTAACAATAGCTTCGATTGGCT 2393

 477 .......PheAsnGlu.............................ThrP 481
           ::::::|||                               |
2394 GAATTTCTACCAAGAAAAATTCGGTAAGAATGATATAAATGGCCTGCTAT 2443

 481 heLeuLysArgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLys 497
     |||||   |||      |||:::::::|||
2444 TTCTGTATCGCTATCGCGATGATGTTCCG..................... 2472

 498 MetValTyrTyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAl 514
               :::   ||||||:::   |||:::   |||      ::
2473 ............ATCGGAAGCTCTAATTTGAAAGAAATGATTATTAGCAG 2510

 514 aThr..ThrLeuAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMe 530
     :   :::|||   :::||||||   :::||||||
2511 TCTTTCTCCCTTGGAGCCTCACTCCCAGAACGAAAT.............TC 2548

 530 tGluValLeuGluAlaLeuCysAspGlySerLeu 541
         ||||||   :::   |||:::::::|||:::
2549 TACAGTATTACTTGTAGCCTGCAACTCCTCAATG 2582
```

OM of: US-09-836-410A-1 to: GenEmbl:* out_format : pfs

Date: Jul 20, 2002 3:43 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09836410/runat_18072002_114913_29213/app_query.fasta_1.660
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836410_@CGN1_1_4586 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-836-410A-1
Query length: 593
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2128.900000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_pat:AX285247	+	3104.00	4364.43	9.7e-235	1779	! AX285247 Sequence 6 from Paten
gb_pat:AX285242	+	3104.00	4358.66	2.0e-234	3418	! AX285242 Sequence 1 from Paten
gb_pat:AX285294	+	3104.00	4358.66	2.0e-234	3418	! AX285294 Sequence 2 from Paten
gb_pat:AX285296	-	3104.00	4358.66	2.0e-234	3418	! AX285296 Sequence 4 from Paten
gb_ro:AF237622	+	3065.00	4303.68	2.4e-231	3421	! AF237622 Mus musculus acetyltr
gb_pr:AF327722	+	3002.50	4213.79	2.4e-226	4192	! AF327722 Homo sapiens putative
gb_pr:HSA314788	+	3002.50	4211.39	3.3e-226	5505	! AJ314788 Homo sapiens mRNA for
gb_ov:AF247679	+	2739.50	3845.13	8.2e-206	3324	! AF247679 Xenopus laevis putati
gb_pat:AX285295	-	1768.00	2483.33	5.8e-130	1413	! AX285295 Sequence 3 from Paten
gb_pr:AK023387	+	1747.00	2451.59	3.4e-128	1802	! AK023387 Homo sapiens cDNA FLJ
gb_pr:AK023402	+	1521.50	2132.89	1.9e-110	1985	! AK023402 Homo sapiens cDNA FLJ
gb_htg:AC015346	+	1385.00	1917.73	1.9e-98	26121	! AC015346 Drosophila melanogas
gb_in:AC010847	+	1385.00	1900.68	1.7e-97	180213	! AC010847 Drosophila melanoga
gb_in:AC011071	+	1385.00	1900.56	1.7e-97	182623	! AC011071 Drosophila melanoga
gb_in:AE003512	+	1385.00	1896.14	3.0e-97	301457	! AE003512 Drosophila melanoga
gb_pr:AB070089	+	1247.00	1746.47	6.5e-89	1876	! AB070089 Macaca fascicularis t
gb_pat:AX210531	-	1176.00	1654.97	8.1e-84	710	! AX210531 Sequence 173 from Pate
gb_pr:BC017099	+	1129.00	1587.21	4.8e-80	843	! BC017099 Homo sapiens, clone IM
gb_pr:AB049844	+	1107.50	1545.94	9.5e-78	2917	! AB049844 Macaca fascicularis b
gb_pl:AY064015	+	989.50	1378.54	2.0e-68	3295	! AY064015 Arabidopsis thaliana
gb_pr:AK025707	+	928.50	1295.35	8.7e-64	2402	! AK025707 Homo sapiens cDNA: FL
gb_pl:SPCC338	+	690.50	935.34	9.8e-44	38734	! AL023781 S.pombe chromosome I
gb_pat:AX245478	+	680.00	960.87	3.7e-45	402	! AX245478 Sequence 408 from Pate
gb_pl:AF263282	+	649.00	867.63	5.8e-40	109900	! AF263282 Filobasidiella neof
gb_pr:AC097376	+	570.00	753.81	1.3e-33	145395	! AC097376 Homo sapiens chromo
gb_pl:YSCNACT	+	550.50	761.53	4.7e-34	2698	! M23166 S.cerevisiae N-acetyltr
gb_pat:I09397	+	550.50	761.44	4.7e-34	2724	! I09397 Sequence 5 from Patent
gb_pl:SCNAT	+	550.50	759.62	6.0e-34	3347	! X15135 Yeast NAT 1 gene for N-
gb_pl:SCYDL040C	-	550.50	759.15	6.4e-34	3530	! Z74088 S.cerevisiae chromosome
gb_pl:SCCIVL37K	-	550.50	738.48	9.0e-33	36687	! Z71781 S.cerevisiae chromosom
gb_pat:I08122	+	546.50	755.89	9.7e-34	2699	! I08122 Sequence 1 from Patent
gb_pl:AC018848	-	478.00	624.92	1.9e-26	132990	! AC018848 Arabidopsis thalian
gb_pl:AC018849	+	473.00	621.34	3.0e-26	89811	! AC018849 Arabidopsis thaliana
gb_pr:AK001595	+	472.00	650.37	7.3e-28	2859	! AK001595 Homo sapiens cDNA FLJ
gb_pat:AX210853	-	452.00	643.52	1.8e-27	255	! AX210853 Sequence 495 from Pate
gb_in:TBR277477	+	442.00	605.57	2.3e-25	3799	! AJ277477 Trypanosoma brucei NA
gb_htg:AC091526	-	430.50	556.17	1.3e-22	163023	! AC091526 Trypanosoma brucei
gb_pr:AL590665	+	430.00	565.10	4.1e-23	54757	! AL590665 Human DNA sequence f
gb_pat:AX245854	+	420.00	594.26	9.7e-25	408	! AX245854 Sequence 784 from Pate
gb_pl:AP002897	-	404.00	519.18	1.5e-20	156425	! AP002897 Oryza sativa genomi
gb_pat:AX245855	+	390.00	551.98	2.2e-22	408	! AX245855 Sequence 785 from P
gb_htg:AC105966	-	383.50	497.85	2.3e-19	66451	! AC105966 Mus musculus clon
gb_pl:SPBC418	+	341.00	461.80	2.3e-17	4459	! AL096876 S.pombe chromosome
gb_htg:AC094440	-	331.00	420.13	4.9e-15	101276	! AC094440 Rattus norvegicu
gb_in:AC024801	+	330.50	417.75	6.6e-15	122450	! AC024801 Caenorhabditis e

seq_name: gb_pat:AX285247

seq_documentation_block:
LOCUS AX285247 1779 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 6 from Patent WO0179506.
ACCESSION AX285247
VERSION AX285247.1 GI:17045931
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Gendron,R.L. and Paradis,H.
TITLE Treatment of ocular neovascularization and related diseases
JOURNAL Patent: WO 0179506-A 6 25-OCT-2001;
Children's Hospital Research Foundation (US)
FEATURES Location/Qualifiers
source 1. .1779
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 627 a 319 c 385 g 448 t
ORIGIN

alignment_scores:
Quality: 3104.00 Length: 593
Ratio: 5.234 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-836-410A-1 x AX285247 ..

Align seg 1/1 to: AX285247 from: 1 to: 1779

```
    1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
      ||||||||||||||||||||||||||||||||||||||||||||||||||
    1 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 50

   17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   51 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGT 100

   34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  101 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 150

   51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 200

   67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 250

   84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  251 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 300

  101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  301 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 350

  117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  351 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 400
```

```
134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 450

151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 500

167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 550

184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 600

201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 650

217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 700

234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 750

251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 800

267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 850

284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 900

301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 950

317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1000

334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1050

351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1100

367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1150

384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1200

401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1250

417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1300

434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1350

451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1400

467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1450

484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1500

501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1550

517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 1600

534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 1650

551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 1700

567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 1750

584 laGluThrGluGluLeuAlaAsnGluIle 593
    |||||||||||||||||||||||||||||
1751 CAGAAACGGAAGAACTGGCCAATGAAATC 1779
```

seq_name: gb_pat:AX285242

```
seq_documentation_block:
LOCUS       AX285242               3418 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 1 from Patent WO0179506.
ACCESSION   AX285242
VERSION     AX285242.1  GI:17045930
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Treatment of ocular neovascularization and related diseases
  JOURNAL   Patent: WO 0179506-A 1 25-OCT-2001;
            Children's Hospital Research Foundation (US)
FEATURES             Location/Qualifiers
     source          1. .3418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1157 a    604 c    704 g    953 t
ORIGIN


alignment_scores:
            Quality: 3104.00              Length:    593
              Ratio:  5.234                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-836-410A-1 x AX285242   ..

 Align seg 1/1  to: AX285242  from: 1  to: 3418
```

```
   1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 408 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 457

  17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 458 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTTATCTGGAGAGAAGT 507

  34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 508 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 557

  51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 558 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 607

  67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 608 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 657

  84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 658 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

 101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 708 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 757

 117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 758 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 807

 134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 808 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 857

 151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 858 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 907

 167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 908 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 957

 184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 958 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 1007

 201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1008 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 1057

 217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1058 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 1107

 234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1108 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 1157

 251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1158 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 1207

 267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1208 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1257

 284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1258 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 1307

 301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1308 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 1357

 317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1358 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1407

 334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1408 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1457

 351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1458 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1507

 367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1508 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1557

 384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1558 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1607

 401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1608 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1657

 417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1658 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1707

 434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1708 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1757

 451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1758 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1807

 467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1808 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1857

 484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1858 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1907

 501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1908 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1957

 517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1958 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 2007

 534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2008 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 2057

 551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2058 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 2107

 567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2108 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 2157

 584 laGluThrGluGluLeuAlaAsnGluIle 593
     |||||||||||||||||||||||||||||
2158 CAGAAACGGAAGAACTGGCCAATGAAATC 2186

seq_name: gb_pat:AX285294
```

```
seq_documentation_block:
LOCUS       AX285294                3418 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 2 from Patent WO0179505.
ACCESSION   AX285294
VERSION     AX285294.1  GI:17045975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 2 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .3418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1157 a    604 c    704 g    953 t
ORIGIN


alignment_scores:
            Quality: 3104.00              Length:    593
              Ratio:  5.234                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-836-410A-1 x AX285294   ..

 Align seg 1/1  to: AX285294  from: 1  to: 3418

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     408 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 457

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     458 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTTATCTGGAGAGAAGT 507

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     508 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 557

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     558 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 607

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     608 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 657

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     658 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     708 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 757

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     758 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 807

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     808 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 857

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     858 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 907

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     908 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 957

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     958 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 1007

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1008 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 1057

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1058 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 1107

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1108 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 1157

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1158 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 1207

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1208 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1257

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1258 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 1307

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1308 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 1357

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1358 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1407

     334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1408 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1457

     351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1458 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1507

     367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1508 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1557

     384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1558 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1607

     401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1608 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1657

     417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1658 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1707

     434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1708 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1757

     451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1758 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1807
```

```
     467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1808 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1857

     484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1858 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1907

     501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1908 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1957

     517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1958 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 2007

     534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2008 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 2057

     551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2058 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 2107

     567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2108 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 2157

     584 laGluThrGluGluLeuAlaAsnGluIle 593
         |||||||||||||||||||||||||||||
    2158 CAGAAACGGAAGAACTGGCCAATGAAATC 2186

seq_name: gb_pat:AX285296

seq_documentation_block:
LOCUS       AX285296               3418 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 4 from Patent WO0179505.
ACCESSION   AX285296
VERSION     AX285296.1  GI:17045977
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 4 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .3418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      953 a    704 c    604 g   1157 t
ORIGIN


alignment_scores:
            Quality: 3104.00              Length:    593
              Ratio:  5.234                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-836-410A-1 x AX285296/rev   ..

 Align seg 1/1  to reverse of: AX285296  from: 1  to: 3418

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    3011 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 2962

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||

    2961 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGT 2912

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2911 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 2862

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2861 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 2812

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2811 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 2762

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2761 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 2712

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2711 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 2662

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2661 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 2612

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2611 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 2562

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2561 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 2512

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2511 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 2462

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2461 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 2412

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2411 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 2362

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2361 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 2312

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2311 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 2262

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2261 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 2212

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2211 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 2162

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2161 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 2112

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2111 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 2062

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    2061 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 2012
```

```
  334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2011 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1962

  351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1961 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1912

  367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1911 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1862

  384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1861 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1812

  401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1811 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1762

  417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1761 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1712

  434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1711 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1662

  451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1661 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1612

  467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1611 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1562

  484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1561 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1512

  501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1511 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1462

  517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1461 TGATGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTGG 1412

  534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1411 AAGCCTTGTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAAGCC 1362

  551 TyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMetPr 567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1361 TACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCATGCC 1312

  567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1311 TCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTTCTG 1262

  584 laGluThrGluGluLeuAlaAsnGluIle 593
      |||||||||||||||||||||||||||||
 1261 CAGAAACGGAAGAACTGGCCAATGAAATC 1233

seq_name: gb_ro:AF237622

seq_documentation_block:
LOCUS       AF237622               3421 bp    mRNA    linear   ROD 13-JUN-2000
DEFINITION  Mus musculus acetyltransferase Tubedown-1 mRNA, complete cds.
ACCESSION   AF237622
VERSION     AF237622.1  GI:8164012
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Tubedown-1, A novel acetyltransferase associated with blood vessel
            development
  JOURNAL   Dev. Dyn. 218 (2), 300-315 (2000)
  MEDLINE   20302532
REFERENCE   2  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Pediatrics, Childrens Hospital Medical
            Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA
REFERENCE   3  (bases 1 to 3421)
  AUTHORS   Gendron,R.L., Adams,L.C. and Paradis,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Pediatrics, Childrens Hospital Medical
            Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA
  REMARK    Amino acid sequence updated by submitter
FEATURES             Location/Qualifiers
     source          1. .3421
                     /organism="Mus musculus"
                     /strain="ICR outbred"
                     /db_xref="taxon:10090"
     CDS             408. .2192
                     /codon_start=1
                     /product="acetyltransferase Tubedown-1"
                     /protein_id="AAF73953.2"
                     /db_xref="GI:8497318"
                     /translation="MLERLKIYEEAWTKYPRGLVPRKLPLNFLSGEKFKECLDRFLRM
                     NFSKGCPPVFNTLRSLYRDKEKVAIVEELVVGYETSLKSCRLFNPNDDGKEEPPTTLL
                     WVQYYLAQHYDKIGQPSIALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMD
                     EAQALDTADRFINSKCAKYMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTE
                     CAQAYKAMNKFGEALKKCHEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVL
                     RQHPFYFKAARIAIEIYLKLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKA
                     QIEEEKKNAEKEKPQRNPKKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPL
                     KNLVKNKIETHLFAFEIYFRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFHSVCES
                     KDLPETVRTVLKQEMNRLFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDSSSQK
                     RAIELATTLDGIPHQQKPSDLHGSVGKPCCDGSLRDCKEAAEAYRASCHKLFPYALAF
                     MPPGYEEDMKITVNGDSSAETEELANEI"
BASE COUNT     1157 a    604 c    706 g    954 t
ORIGIN

alignment_scores:
            Quality: 3065.00              Length:    596
              Ratio:  5.169                 Gaps:      3
 Percent Similarity:  99.497   Percent Identity: 99.497

alignment_block:
 US-09-836-410A-1 x AF237622   ..

 Align seg 1/1  to: AF237622  from: 1  to: 3421

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     408 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 457

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     458 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGT 507

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     508 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 557

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     558 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 607

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 608 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 657

  84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 658 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

 101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 708 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 757

 117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 758 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 807

 134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 808 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 857

 151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 858 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 907

 167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 908 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 957

 184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 958 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 1007

 201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1008 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 1057

 217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1058 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 1107

 234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1108 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 1157

 251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1158 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 1207

 267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1208 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1257

 284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1258 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 1307

 301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1308 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 1357

 317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1358 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1407

 334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1408 CCCAGATTGAAGAAGAGAAAAAAAATGCCGAAAAAGAAAAGCCGCAACGG 1457

 351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1458 AATCCGAAAAAGAAAAAGGATGATGATGACGAAGAAATTGGAGGCCCCAA 1507

 367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1508 AGAAGAGCTTATCCCTGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 1557

 384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1558 AAGCTATTAAGTTTTTAACACCATTGAAGAACTTGGTGAAGAACAAGATA 1607

 401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1608 GAAACTCATCTTTTTGCCTTTGAGATCTACTTTAGGAAAGAAAAGTTTCT 1657

 417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1658 TTTGATGCTACAATCAGTAAAGCGGGCATTTGCTATTGATTCTAGTCATC 1707

 434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1708 CCTGGCTTCATGAGTGCATGATTCGACTCTTTCATTCTGTGTGTGAAAGT 1757

 451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1758 AAAGACTTACCCGAAACAGTTAGAACAGTATTAAAACAAGAAATGAATCG 1807

 467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1808 TCTTTTTGGAGCAACAAATCCAAAGAATTTTAATGAAACCTTTCTGAAAA 1857

 484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1858 GGAATTCTGATTCATTGCCACATAGATTATCAGCTGCCAAAATGGTATAT 1907

 501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1908 TATTTAGATTCTTCTAGTCAAAAACGAGCAATAGAGCTGGCGACAACACT 1957

 517 uAsp.GlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeu 533
     |||| |||||||||||||||||||||||||||||||||||||||||||||
1958 TGATGGGATCCCTCACCAACAGAAACCTTCAGACTTGCATGGAAGTGTTG 2007

 534 .GluAlaLeu.CysAspGlySerLeuArgAspCysLysGluAlaAlaGlu 549
      ||||||||| |||||||||||||||||||||||||||||||||||||||
2008 GGAAGCCTTGTTGTGATGGTAGCCTACGAGACTGTAAAGAAGCTGCCGAA 2057

 550 AlaTyrArgAlaSerCysHisLysLeuPheProTyrAlaLeuAlaPheMe 566
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2058 GCCTACAGAGCAAGTTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCAT 2107

 566 tProProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerS 583
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2108 GCCTCCTGGATACGAAGAGGATATGAAGATCACAGTGAACGGAGATAGTT 2157

 583 erAlaGluThrGluGluLeuAlaAsnGluIle 593
     ||||||||||||||||||||||||||||||||
2158 CTGCAGAAACGGAAGAACTGGCCAATGAAATC 2189
```

seq_name:!gb_pr:AF327722.

```
seq_documentation_block:
LOCUS       AF327722               4192 bp    mRNA    linear   PRI 05-MAR-2001
DEFINITION  Homo sapiens putative acetyltransferase mRNA, complete cds.
ACCESSION   AF327722
VERSION     AF327722.1  GI:13195459
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4192)
  AUTHORS   He,Y.G., Tan,D.Y., Lai,J.H., Xie,Y.F. and Qian,W.
  TITLE     Cloning and analysis of a human putative acetyltransferase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4192)
  AUTHORS   He,Y.G., Tan,D.Y., Lai,J.H., Xie,Y.F. and Qian,W.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (11-DEC-2000) Biology Department, Yunnan University,
            North Street of Greenlake, Kunming, Yunnan 650091, China
FEATURES             Location/Qualifiers
     source          1. .4192
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             280. .2880
                     /codon_start=1
                     /product="putative acetyltransferase"
                     /protein_id="AAK15707.1"
                     /db_xref="GI:13195460"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                     LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                     QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                     LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                     DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                     ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                     YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                     HEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVLRQHPFYFKAARIAIEIYL
                     KLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKAQIEEEKKNAEKEKQQRNQ
                     KKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPLKNLVKNKIETHLFAFEIY
                     FRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFNTAVCESKDLSDTVRTVLKQEMNR
                     LFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDPSSQKRAIELATTLDESLTNRN
                     LQTCMEVLEALYDGSLGDCKEAAEIYRANCHKLFPYALAFMPPGYEEDMKITVNGDSS
                     AEAEELANEI"
BASE COUNT     1427 a    710 c    953 g    1102 t
ORIGIN


alignment_scores:
            Quality: 3002.50              Length:    594
              Ratio:  5.132               Gaps:        1
 Percent Similarity:  98.485   Percent Identity: 96.970

alignment_block:
 US-09-836-410A-1 x AF327722   ..

 Align seq 1/1  to: AF327722  from: 1  to: 4192

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1096 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTGGACTAAATATCCCAG 1145

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||:::||||||||||||||||||||||||||||||
    1146 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTATCTGGTGAGAAGT 1195

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
    1196 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 1245

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
    1246 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 1295

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         |||||||:::||||||||||||||||||||||||||||||||||||||||
    1296 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1345

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1346 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1395

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1396 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1445

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1446 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1495

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1496 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1545

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1546 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1595

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1596 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1645

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1646 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1695

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1696 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1745

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1746 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1795

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1796 AGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAG 1845

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1846 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1895

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1896 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1945

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1946 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCC 1995

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1996 CTTACAGATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGA 2045

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2046 CAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAG 2095

     334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2096 CCCAGATAGAAGAAGAGAAAAAAAATGCAGAAAAAGAAAAGCAGCAGAGA 2145

     351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2146 AATCAGAAAAAGAAGAAGGATGATGATGATGAGGAGATAGGAGGTCCAAA 2195

     367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2196 AGAAGAACTTATTCCAGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 2245

     384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2246 AAGCTATTAAATTTTTAACACCGTTGAAGAACTTGGTGAAGAACAAGATA 2295

     401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2296 GAGACTCATCTTTTTGCCTTTGAGATTTACTTTAGGAAAGAAAAGTTTCT 2345

     417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2346 TTTGATGCTACAATCAGTAAAGAGGGCATTTGCTATTGATTCTAGTCATC 2395

     434 roTrpLeuHisGluCysMetIleArgLeuPheHis...SerValCysGlu 449
```

```
       ||||||||||||||||||||||||||||||||:::    :::|||||||||
2396 CCTGGCTTCATGAGTGTATGATTCGTCTCTTTAATACTGCAGTGTGTGAA 2445

 450 SerLysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAs 466
     |||||||||||    :::||||||||||||||||||||||||||||||||
2446 AGTAAAGATTTATCTGATACAGTTAGAACAGTATTAAAACAAGAAATGAA 2495

 466 nArgLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuL 483
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2496 TCGTCTTTTTGGAGCAACGAATCCAAAGAATTTTAATGAAACTTTTCTGA 2545

 483 ysArgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetVal 499
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2546 AAAGGAATTCTGATTCATTGCCACACAGATTATCAGCTGCCAAAATGGTA 2595

 500 TyrTyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrTh 516
     |||||||||||   ||||||||||||||||||||||||||||||||||||
2596 TATTACTTAGATCCTTCTAGTCAGAAGCGAGCTATAGAGTTGGCAACAAC 2645

 516 rLeuAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValL 533
     ||||||   |||||||||||||||||||||||||||||||||||||||||
2646 ACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGGAGGTAT 2695

 533 euGluAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGlu 549
     |||||||||||    |||||||||||    ||||||||||||||||||||
2696 TGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAA 2745

 550 AlaTyrArgAlaSerCysHisLysLeuPheProTyrAlaLeuAlaPheMe 566
        |||||||||:::||||||||||||||||||||||||||||||||||||
2746 ATTTATAGAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCAT 2795

 566 tProProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerS 583
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2796 GCCTCCTGGATATGAAGAGGATATGAAGATCACAGTTAATGGAGATAGTT 2845

 583 erAlaGluThrGluGluLeuAlaAsnGluIle 593
     ||||||||:::||||||||||||||||||||
2846 CTGCAGAAGCTGAAGAACTGGCCAATGAAATT 2877
```

seq_name: gb_pr:HSA314788

```
seq_documentation_block:
LOCUS       HSA314788            5505 bp    mRNA    linear   PRI 01-JUL-2001
DEFINITION  Homo sapiens mRNA for putative N-acetyltransferase.
ACCESSION   AJ314788
VERSION     AJ314788.1  GI:14589341
KEYWORDS    N-acetyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5505)
  AUTHORS   Fluge,O., Bruland,O., Akslen,L.A., Varhaug,J.E. and Lillehaug,J.R.
  TITLE     Identification of NATH, a novel gene overexpressed in papillary
            thyroid carcinomas
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5505)
  AUTHORS   Fluge,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2001) Fluge O., Dept. of Molecular Biology,
            University of Bergen, Thormohlens gt 55, N-5020 Bergen, NORWAY
COMMENT     related entry AF327722.
FEATURES             Location/Qualifiers
     source          1..5505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="papillary thyroid carcinoma"
     CDS             209..2809
                     /codon_start=1
                     /evidence=experimental
                     /product="putative N-acetyltransferase"
                     /protein_id="CAC43228.1"
                     /db_xref="GI:14589342"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                     LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                     QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                     LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                     DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                     ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                     YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                     HEIERHFIEITDDQFDFHTYCMRKITLRSYVDLLKLEDVLRQHPFYFKAARIAIEIYL
                     KLHDNPLTDENKEHEADTANMSDKELKKLRNKQRRAQKKAQIEEEKKNAEKEKQQRNQ
                     KKKKDDDDEEIGGPKEELIPEKLAKVETPLEEAIKFLTPLKNLVKNKIETHLFAFEIY
                     FRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFNTAVCESKDLSDTVRTVLKQEMNR
                     LFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYYLDPSSQKRAIELATTLDESLTNRN
                     LQTCMEVLEALYDGSLGDCKEAAEIYRANCHKLFPYALAFMPPGYEEDMKITVNGDSS
                     AEAEELANEI"
     conflict        3752..5505
                     /citation=[1]
BASE COUNT     1847 a    914 c   1154 g   1590 t
ORIGIN
```

```
alignment_scores:
        Quality:  3002.50          Length:     594
          Ratio:  5.132              Gaps:       1
 Percent Similarity:  98.485   Percent Identity:  96.970

alignment_block:
 US-09-836-410A-1 x HSA314788  ..

 Align seg 1/1  to: HSA314788  from: 1  to: 5505

   1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1025 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTGGACTAAATATCCCAG 1074

  17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
1075 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 1124

  34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
1125 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 1174

  51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
     |||||||||||||||||||||||||||||||||:::||||||||||||||
1175 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 1224

  67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
     ||||||:::|||||||||||||||||||||||||||||||||||||||||
1225 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1274

  84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1275 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1324

 101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1325 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1374

 117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1375 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1424

 134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1425 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1474

 151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1475 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1524

 167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
```

```
1525 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1574

 184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
1575 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1624

 201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
1625 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1674

 217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
1675 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1724

 234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
1725 AGAAATGTCATGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAG 1774

 251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
1775 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1824

 267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
1825 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1874

 284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
1875 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCC 1924

 301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
1925 CTTACAGATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGA 1974

 317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
1975 CAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAG 2024

 334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
2025 CCCAGATAGAAGAAGAGAAAAAAAATGCAGAAAAAGAAAAGCAGCAGAGA 2074

 351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
2075 AATCAGAAAAAGAAGAAGGATGATGATGATGAGGAGATAGGAGGTCCAAA 2124

 367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
2125 AGAAGAACTTATTCCAGAGAAACTGGCCAAGGTTGAAACTCCATTGGAAG 2174

 384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
2175 AAGCTATTAAATTTTTAACACCGTTGAAGAACTTGGTGAAGAACAAGATA 2224

 401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
2225 GAGACTCATCTTTTTGCCTTTGAGATTTACTTTAGGAAAGAAAAGTTTCT 2274

 417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
2275 TTTGATGCTACAATCAGTAAAGAGGGCATTTGCTATTGATTCTAGTCATC 2324

 434 roTrpLeuHisGluCysMetIleArgLeuPheHis...SerValCysGlu 449
2325 CCTGGCTTCATGAGTGTATGATTCGTCTCTTTAATACTGCAGTGTGTGAA 2374

 450 SerLysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAs 466
2375 AGTAAAGATTTATCTGATACAGTTAGAACAGTATTAAAACAAGAAATGAA 2424

 466 nArgLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuL 483
2425 TCGTCTTTTTGGAGCAACGAATCCAAAGAATTTTAATGAAACTTTTCTGA 2474

 483 ysArgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetVal 499
2475 AAAGGAATTCTGATTCATTGCCACACAGATTATCAGCTGCCAAAATGGTA 2524

 500 TyrTyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrTh 516
2525 TATTACTTAGATCCTTCTAGTCAGAAGCGAGCTATAGAGTTGGCAACAAC 2574

 516 rLeuAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValL 533
2575 ACTTGATGAATCTCTCACTAACAGAAACCTCCAGACATGTATGGAGGTAT 2624

 533 euGluAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGlu 549
2625 TGGAAGCCTTGTATGATGGTAGCCTAGGAGACTGTAAAGAAGCTGCTGAA 2674

 550 AlaTyrArgAlaSerCysHisLysLeuProTyrAlaLeuAlaPheMe 566
2675 ATTTATAGAGCAAATTGTCATAAGCTTTTCCCTTATGCTTTGGCTTTCAT 2724

 566 tProProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerS 583
2725 GCCTCCTGGATATGAAGAGGATATGAAGATCACAGTTAATGGAGATAGTT 2774

 583 erAlaGluThrGluGluLeuAlaAsnGluIle 593
2775 CTGCAGAAGCTGAAGAACTGGCCAATGAAATT 2806
```

seq_name: gb_ov:AF247679

seq_documentation_block:

```
LOCUS       AF247679               3324 bp    mRNA    linear   VRT 02-AUG-2000
DEFINITION  Xenopus laevis putative N-terminal acetyltransferase mRNA, complete
            cds.
ACCESSION   AF247679
VERSION     AF247679.1  GI:9651962
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 3324)
  AUTHORS   Choi,S.-C., Kim,J. and Han,J.-K.
  TITLE     Expression of N-terminal acetyltransferase in Xenopus laevis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3324)
  AUTHORS   Choi,S.-C., Kim,J. and Han,J.-K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Life Science, Pohang University of Science
            and Technology, Pohang 790-784, South Korea
FEATURES             Location/Qualifiers
     source          1. .3324
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     CDS             105. .2645
                     /codon_start=1
                     /product="putative N-terminal acetyltransferase"
                     /protein_id="AAF91333.1"
                     /db_xref="GI:9651963"
                     /translation="MPSVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEDAYDLVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQLRPAQRASWIGYAIAYHLLEDYEMAAKILEEFRKTQ
                     QTSPDKVDYEYSELLLYQNQVLREAELYKEALDHLCTYEKQICDKLAVEETKGDLLNR
                     LGRLDEAADVYRALQDRCPENWAYYKGLEKALKPESMEERLKIYEDSWTKYPRGLVSR
                     RLPLNFLSGLKFRECLDKYLRMNFSKGCPPVFNTLRSLYSDKEKVEIIEDLVVGYETS
                     LKSCRLFNMNDDGKEEPPTTLLWVQYYLAQHYDKIGQQSIALEYINAAIDSTPTLIEL
                     FLVKAKIYKHSGNIKEAVRWMDEAQALDTADRFINSKCAKYMLKANLIKEAEEMCSKF
                     TREGTSAVENLNEMQCMWFQTECAQAYKSMNKYGEALKKCHEIERHFVEITDDQFDFH
                     TYCMRKITLRSYADLLKLEDVLRQHPFYFKAARIAIEIYLKLHDNPLTDENKEHEADT
                     ANMSDKELKKLRNKQRRAQKKAQLEEEKKNAEKEKQQRNQKKKKEDDDEEIGGPKEEL
```

```
                    IPEKLAKVENPLEEAIKFLTPLKNLVKNKIETHLYAFEIYFRKDKFLLMLQSVKRAYA
                    IDPNHPWLHQCLIRFFCAVSESKELNESVRTVLKQEMCRLFGETSPANFNNSFLKENI
                    NSIPHRLAAAKMMYYLDHSSQKRSVELGTSLDESLCNGSLQTCTDVLEALRDGSLGDK
                    EAAECYRVSCHKLYPYALAFMPPGYEEDMSITMNGDSSAETEELANEM"
BASE COUNT     1111 a    599 c    756 g    858 t
ORIGIN


alignment_scores:
            Quality: 2739.50              Length:    593
              Ratio:  4.823                Gaps:      1
 Percent Similarity:  95.784   Percent Identity: 87.521

alignment_block:
 US-09-836-410A-1 x AF247679   ..

 Align seg 1/1  to: AF247679  from: 1  to: 3324

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         |||   |||||||||||||||||||:::::|||||||||||||||||||
     867 ATGGAAGAAAGACTGAAAATTTATGAAGACAGCTGGACAAAGTATCCCAG 916

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||   |||:::|||||||||||||||||||||   ||||
     917 GGGATTAGTTTCAAGAAGACTGCCTTTAAACTTCTTATCAGGTCTAAAAT 966

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||:::|||||||||||:::::::|||||||||||||||||||||||||||
     967 TTAGAGAATGTCTGGATAAATATTTGAGGATGAATTTTAGCAAAGGATGC 1016

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         |||||||||||||||||||||||||||||||   |||||||||||||||
    1017 CCTCCAGTATTTAATACATTGAGATCACTCTACAGCGATAAAGAGAAGGT 1066

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         |   |||:::|||:::|||||||||||||||||||||||||||||||||||
    1067 AGAGATTATTGAAGATTTAGTTGTTGGTTATGAAACATCCTTGAAAAGCT 1116

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||   ||||||||||||||||||||||||||||||||||
    1117 GCCGTTTATTTAACATGAATGATGATGGGAAAGAAGAACCACCAACAACT 1166

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1167 TTACTGTGGGTGCAGTACTACTTGGCACAACATTATGACAAAATTGGCCA 1216

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         |   |||||||||||||||||||||||||:::|||||:::||||||||||
    1217 GCAATCAATTGCTCTGGAATATATTAATGCTGCGATTGACAGCACACCAA 1266

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
    1267 CTCTTATTGAACTTTTCCTTGTCAAAGCTAAAATTTACAAGCATTCTGGG 1316

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
    1317 AATATAAAAGAAGCTGTCCGATGGATGGATGAAGCGCAGGCACTGGATAC 1366

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1367 AGCTGACCGATTTATCAATTCTAAATGTGCAAAATATATGTTAAAAGCAA 1416

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1417 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCGAAATTTACAAGGGAAGGA 1466

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1467 ACATCAGCAGTGGAAAATCTGAATGAGATGCAGTGCATGTGGTTCCAGAC 1516

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||:::|||||||||:::|||||||||||||||
    1517 AGAATGTGCACAAGCTTACAAGTCCATGAATAAATATGGTGAAGCACTTA 1566

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||:::|||||||||||||||||
    1567 AAAAATGCCATGAAATTGAAAGACATTTTGTAGAAATTACAGATGACCAG 1616

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||::
    1617 TTTGATTTCCACACTTACTGTATGAGAAAAATTACACTCAGGTCATATGC 1666

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         :|||||||||||||||||||||||||||||||||||||||||||||||||
    1667 GGACTTATTAAAATTGGAGGATGTACTAAGGCAGCATCCATTTTACTTCA 1716

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1717 AGGCTGCACGGATTGCAATAGAAATTTATCTGAAGCTTCATGATAATCCT 1766

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1767 CTAACAGATGAAAATAAAGAGCATGAGGCAGACACAGCCAATATGTCTGA 1816

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1817 CAAAGAGCTGAAGAAACTACGTAACAAACAACGGAGGGCGCAGAAGAAAG 1866

     334 laGlnIleGluGluGluLysLysAsnAlaGluLysGluLysProGlnArg 350
         |||||:::|||||||||||||||||||||||||||||||   ||||||
    1867 CACAGCTGGAAGAGGAGAAAAAGAACGCGGAAAAAGAAAAGCAACAGAGA 1916

     351 AsnProLysLysLysLysAspAspAspAspGluGluIleGlyGlyProLy 367
         |||   ||||||||||||:::|||||||||||||||||||||||||||||
    1917 AACCAAAAGAAGAAAAAAGAAGATGATGATGAAGAGATTGGAGGACCGAA 1966

     367 sGluGluLeuIleProGluLysLeuAlaLysValGluThrProLeuGluG 384
         ||||||||||||||||||||||||||||||||||||:::||||||||||||
    1967 GGAAGAGCTTATACCAGAAAAGTTAGCAAAGGTTGAAAATCCTTTGGAAG 2016

     384 luAlaIleLysPheLeuThrProLeuLysAsnLeuValLysAsnLysIle 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2017 AAGCTATTAAATTTTTAACACCTCTAAAGAACTTGGTGAAGAACAAAATA 2066

     401 GluThrHisLeuPheAlaPheGluIleTyrPheArgLysGluLysPheLe 417
         ||||||||||||:::|||||||||||||||||||||||||:::|||||||
    2067 GAGACACATCTTTATGCCTTTGAAATTTACTTCAGGAAAGATAAGTTTCT 2116

     417 uLeuMetLeuGlnSerValLysArgAlaPheAlaIleAspSerSerHisP 434
         ||||||||||||||||||||||||||||:::|||||||||   :::||||
    2117 TCTGATGTTGCAGTCTGTGAAACGAGCATACGCAATAGATCCTAACCACC 2166

     434 roTrpLeuHisGluCysMetIleArgLeuPheHisSerValCysGluSer 450
         ||||||||||:::|||:::|||||:::|||   :::|||   ||||||
    2167 CGTGGCTTCATCAGTGTCTTATCCGATTCTTTTGTGCAGTTTCTGAAAGT 2216

     451 LysAspLeuProGluThrValArgThrValLeuLysGlnGluMetAsnAr 467
         |||:::|||   |||:::|||||||||||||||||||||||||   ||
    2217 AAAGAACTAAACGAATCGGTAAGAACAGTGTTAAAGCAAGAAATGTGCAG 2266

     467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
         ||||||||||   |||:::|||   |||||||||:::::|||||||||:
    2267 ACTTTTTGGAGAAACCAGCCCTGCAAATTTTAATAATTCCTTCCTGAAAG 2316

     484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
         ::|||   :::|||:::||||||||||||:::|||||||||||||:::|||
    2317 AAAACATCAACTCCATACCACACCGATTAGCAGCTGCCAAAATGATGTAC 2366

     501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
         |||||||||   |||||||||||||||:::::|||||:::|||:::|||
    2367 TATCTGGACCATTCAAGTCAAAAACGATCAGTAGAGCTGGGGACCTCATT 2416

     517 uAspGlySerLeuThrAsnArgAsnLeuGlnThrCysMetGluValLeuG 534
         ||||   ||||||   |||   :::|||||||||||||   :::|||||||
    2417 GGATGAATCGCTCTGTAACGGTAGTCTCCAGACGTGCACCGATGTTCTGG 2466
```

```
    534 luAlaLeuCysAspGlySerLeuArgAspCysLysGluAlaAlaGluAla 550
        ||||||||   ||||||||||||   |||   ||||||||||||||:::
   2467 AAGCCTTGCGAGACGGCAGTTTAGGAGAC...AAAGAGGCTGCTGAATGT 2513

    551 TyrArgAlaSerCysHisLysLeuPheProTyrAlaLeuAlaPheMetPr 567
        ||||||:::|||||||||||||||:::|||||||||||||||||||||||
   2514 TATCGAGTCAGTTGCCACAAGCTCTACCCTTACGCTTTAGCTTTCATGCC 2563

    567 oProGlyTyrGluGluAspMetLysIleThrValAsnGlyAspSerSerA 584
        ||||||||||||||||||||||:::||||||:::|||||||||||||||||
   2564 TCCTGGCTACGAAGAGGACATGAGTATCACAATGAATGGAGATAGCTCCG 2613

    584 laGluThrGluGluLeuAlaAsnGluIle 593
        |||||||||||||||||||||||||:::
   2614 CCGAAACAGAAGAACTGGCTAATGAAATG 2642

seq_name: gb_pat:AX285295

seq_documentation_block:
LOCUS       AX285295               1413 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 3 from Patent WO0179505.
ACCESSION   AX285295
VERSION     AX285295.1  GI:17045976
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Gendron,R.L. and Paradis,H.
  TITLE     Inhibition of bone tumor formation using antisense cdna therapy
  JOURNAL   Patent: WO 0179505-A 3 25-OCT-2001;
            CHILDREN'S HOSPITAL MEDICAL CENTER (US)
FEATURES             Location/Qualifiers
     source          1. .1413
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      359 a    314 c    244 g    496 t
ORIGIN


alignment_scores:
            Quality: 1768.00              Length:    335
              Ratio:  5.278                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-836-410A-1 x AX285295/rev   ..

 Align seg 1/1  to reverse of: AX285295  from: 1  to: 1413

       1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1006 ATGTTAGAACGGCTAAAAATATATGAGGAAGCCTGGACTAAATACCCCAG 957

      17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     956 GGGACTCGTGCCAAGAAAGCTGCCCTTAAACTTTTATCTGGAGAGAAGT 907

      34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     906 TTAAGGAGTGTTTGGATAGGTTCCTAAGGATGAATTTCAGCAAGGGCTGT 857

      51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     856 CCACCTGTCTTCAATACCTTGAGGTCTTTATACAGAGATAAAGAGAAGGT 807

      67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     806 GGCAATCGTAGAAGAACTAGTAGTTGGTTATGAAACTTCTCTAAAAAGTT 757

      84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     756 GTCGCCTATTTAACCCCAATGATGATGGAAAGGAGGAACCTCCAACCACA 707

     101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     706 TTACTTTGGGTCCAGTACTATTTGGCACAGCATTATGATAAAATTGGTCA 657

     117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     656 GCCATCCATTGCTCTGGAATACATAAATACTGCAATTGAAAGTACACCAA 607

     134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     606 CATTGATAGAACTCTTTCTTGTAAAAGCTAAAATCTATAAGCATGCTGGG 557

     151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     556 AATATTAAAGAAGCTGCCAGGTGGATGGATGAAGCCCAGGCCCTGGACAC 507

     167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     506 AGCAGACAGATTTATTAATTCCAAGTGTGCAAAATACATGTTAAAAGCCA 457

     184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     456 ACCTGATTAAAGAGGCTGAAGAAATGTGTTCCAAGTTTACGAGGGAAGGA 407

     201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     406 ACTTCAGCGGTAGAGAACCTGAATGAAATGCAGTGTATGTGGTTCCAGAC 357

     217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     356 AGAGTGTGCTCAGGCATACAAAGCAATGAACAAATTTGGTGAAGCACTTA 307

     234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     306 AGAAATGTCATGAAATTGAGAGACATTTTATAGAAATCACCGATGACCAG 257

     251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 TTTGACTTTCATACATACTGTATGAGGAAGATCACCCTTAGATCATATGT 207

     267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 157

     284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     156 AAGCAGCGAGAATTGCTATTGAGATCTATTTGAAGCTTCATGACAACCCT 107

     301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     106 CTGACAGATGAGAACAAAGAACACGAGGCTGATACAGCAAACATGTCTGA 57

     317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      56 CAAAGAGCTAAAGAAACTGCGTAATAAACAAAGAAGAGCTCAAAAGAAAG 7

     334 laGln 335
         |||||
       6 CCCAG 2

seq_name: gb_pr:AK023387

seq_documentation_block:
LOCUS       AK023387               1802 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ13325 fis, clone OVARC1001762, weakly similar
            to N-TERMINAL ACETYLTRANSFERASE 1 (EC 2.3.1.88).
ACCESSION   AK023387
VERSION     AK023387.1  GI:10435302
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_lib:OVARC1
            clone:OVARC1001762.
```

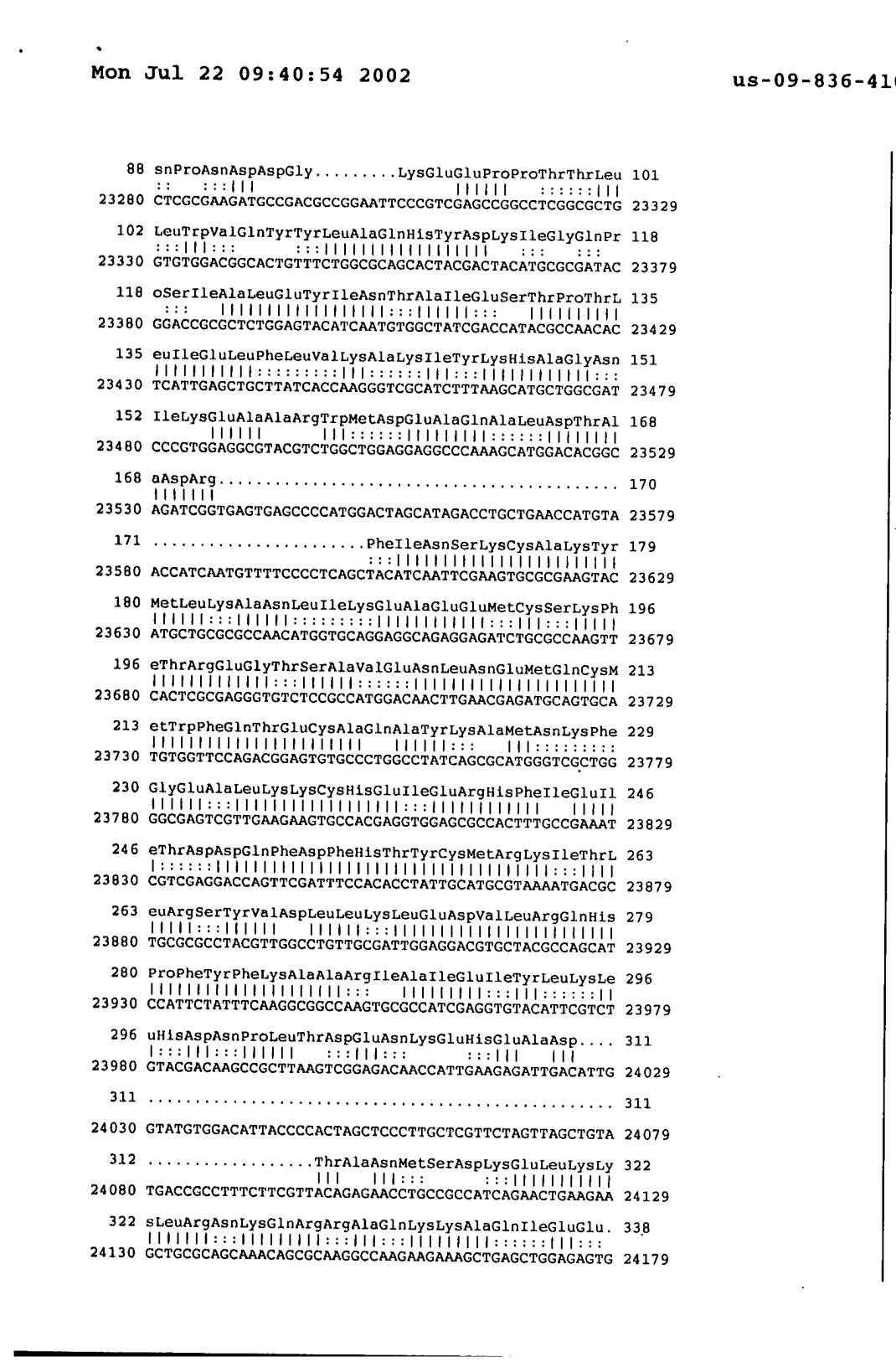
```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1802)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="OVARC1001762"
                     /clone_lib="OVARC1"
                     /tissue_type="ovary, tumor tissue"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      644 a    298 c    395 g    465 t
ORIGIN


alignment_scores:
           Quality: 1747.00              Length:    336
             Ratio:  5.215                Gaps:      0
 Percent Similarity:  99.702   Percent Identity: 98.512

alignment_block:
 US-09-836-410A-1 x AK023387   ..

 Align seg 1/1  to: AK023387  from: 1  to: 1802

      1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
        |||||||||||||||||||||||||||||||||   |||||||||||||||
    794 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTAGACTAAATATCCCAG 843

     17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
        ||||||||||||||||:::||||||||||||||||||||||||||||||
    844 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 893

     34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
        ||||||||||||||||:::||||||||||||||||||||||||||||||
    894 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 943

     51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    944 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 993

     67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
        |||||||:::||||||||||||||||||||||||||||||||||||||||
    994 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1043

     84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1044 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1093

    101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1094 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1143

    117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1144 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1193

    134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1194 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1243

    151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1244 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1293

    167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1294 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1343

    184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1344 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1393

    201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1394 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1443

    217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1444 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1493

    234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1494 AGAAATGTCACGAGATTGAGAGACATTTTATAGAAATCACTGATGACCAG 1543

    251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1544 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1593

    267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1594 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1643

    284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAspAsnPro 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1644 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGACAACCCC 1693

    301 LeuThrAspGluAsnLysGluHisGluAlaAspThrAlaAsnMetSerAs 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1694 CTTACAGATGAGAATAAAGAACACGAAGCTGATACAGCAAACATGTCTGA 1743

    317 pLysGluLeuLysLysLeuArgAsnLysGlnArgArgAlaGlnLysLysA 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1744 CAAAGAGCTAAAGAAGCTACGTAATAAACAAAGAAGAGCTCAAAAGAAAG 1793

    334 laGlnIle 336
        |||||||
   1794 CCCAGATA 1801

seq_name: gb_pr:AK023402

seq_documentation_block:
LOCUS       AK023402               1985 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ13340 fis, clone OVARC1001942, weakly similar
            to N-TERMINAL ACETYLTRANSFERASE 1 (EC 2.3.1.88).
ACCESSION   AK023402
VERSION     AK023402.1  GI:10435324
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_lib:OVARC1
            clone:OVARC1001942.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1985)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1985
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="OVARC1001942"
                     /tissue_type="ovary, tumor tissue"
                     /clone_lib="OVARC1"
                     /note="cloning vector: pME18SFL3"
     CDS             280. .1860
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB14562.1"
                     /db_xref="GI:10435325"
                     /translation="MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAE
                     HGETLAMKGLTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKC
                     YRNALKWDKDNLQILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHL
                     LEDYEMAAKILEEFRKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEK
                     QICDKLAVEETKGELLLQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLER
                     LKIYEEAWTKYPRGLVPRRLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYK
                     DKEKVAIIEELVVGYETSLKSCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSI
                     ALEYINTAIESTPTLIELFLVKAKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAK
                     YMLKANLIKEAEEMCSKFTREGTSAVENLNEMQCMWFQTECAQAYKAMNKFGEALKKC
                     HEIERKSLMTSLTFIHTV"
BASE COUNT      652 a    349 c    492 g    492 t
ORIGIN


alignment_scores:
            Quality: 1521.50              Length:    298
              Ratio:  5.175                Gaps:      1
 Percent Similarity:  98.658   Percent Identity: 97.315

alignment_block:
 US-09-836-410A-1 x AK023402   ..

 Align seg 1/1  to: AK023402  from: 1  to: 1985

      1 MetLeuGluArgLeuLysIleTyrGluGluAlaTrpThrLysTyrProAr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1096 ATGTTAGAACGGCTAAAAATTTATGAGGAAGCCTGGACTAAATATCCCAG 1145

     17 gGlyLeuValProArgLysLeuProLeuAsnPheLeuSerGlyGluLysP 34
        |||||||||||||||:::|||||||||||||||||||||||||||||||
   1146 GGGACTGGTGCCAAGAAGGCTGCCGTTAAACTTTTTATCTGGTGAGAAGT 1195

     34 heLysGluCysLeuAspArgPheLeuArgMetAsnPheSerLysGlyCys 50
        |||||||||||||||:::|||||||||||||||||||||||||||||||
   1196 TTAAAGAATGTTTGGATAAGTTCCTAAGGATGAATTTCAGCAAGGGTTGC 1245

     51 ProProValPheAsnThrLeuArgSerLeuTyrArgAspLysGluLysVa 67
        ||||||||||||||||||||||||||||||||:::|||||||||||||||
   1246 CCACCAGTCTTCAATACTTTAAGATCATTATACAAAGACAAAGAAAAGGT 1295

     67 lAlaIleValGluGluLeuValValGlyTyrGluThrSerLeuLysSerC 84
        |||||||:::|||||||||||||||||||||||||||||||||||||||
   1296 GGCAATCATAGAAGAGTTAGTAGTAGGTTATGAAACCTCTCTAAAAAGCT 1345

     84 ysArgLeuPheAsnProAsnAspAspGlyLysGluGluProProThrThr 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1346 GCCGGTTATTTAACCCCAATGATGATGGAAAGGAGGAACCACCAACCACA 1395

    101 LeuLeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1396 TTACTTTGGGTCCAGTACTACTTGGCACAACATTATGACAAAATTGGTCA 1445

    117 nProSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1446 GCCATCTATTGCTTTGGAGTACATAAATACTGCTATTGAAAGTACACCTA 1495

    134 hrLeuIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGly 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1496 CATTAATAGAACTCTTTCTCGTGAAAGCTAAAATCTATAAGCATGCTGGA 1545

    151 AsnIleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1546 AATATTAAAGAAGCTGCAAGGTGGATGGATGAGGCCCAGGCCTTGGACAC 1595

    167 rAlaAspArgPheIleAsnSerLysCysAlaLysTyrMetLeuLysAlaA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1596 AGCAGACAGATTTATCAACTCCAAATGTGCAAAATACATGCTAAAAGCCA 1645

    184 snLeuIleLysGluAlaGluGluMetCysSerLysPheThrArgGluGly 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1646 ACCTGATTAAAGAAGCTGAAGAAATGTGCTCAAAGTTTACAAGGGAAGGA 1695

    201 ThrSerAlaValGluAsnLeuAsnGluMetGlnCysMetTrpPheGlnTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1696 ACATCAGCGGTAGAGAATTTGAATGAAATGCAGTGCATGTGGTTCCAAAC 1745

    217 rGluCysAlaGlnAlaTyrLysAlaMetAsnLysPheGlyGluAlaLeuL 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1746 AGAATGTGCCCAGGCTTATAAAGCAATGAATAAATTTGGTGAAGCACTTA 1795

    234 ysLysCysHisGluIleGluArgHisPheIleGluIleThrAspAspGln 250
        |||||||||||||||||||||          ||||||||||||||||
   1796 AGAAATGTCATGAGATTGAGAGAAA.........ATCACTGATGACCAG 1835

    251 PheAspPheHisThrTyrCysMetArgLysIleThrLeuArgSerTyrVa 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1836 TTTGACTTTCATACATACTGTATGAGGAAGATTACCCTTAGATCATATGT 1885

    267 lAspLeuLeuLysLeuGluAspValLeuArgGlnHisProPheTyrPheL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1886 GGACTTATTAAAACTAGAAGATGTACTTCGACAGCATCCATTTTACTTCA 1935

    284 ysAlaAlaArgIleAlaIleGluIleTyrLeuLysLeuHisAsp 298
        |||||||||||||||||||||||||||||||||||||||||||
   1936 AGGCAGCAAGAATTGCTATAGAGATCTATTTGAAGCTTCATGAC 1979

seq_name: gb_htg:AC015346

seq_documentation_block:
LOCUS       AC015346              26121 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015346
VERSION     AC015346.1  GI:6435989
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
```

```
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 26121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210169 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .26121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     6182 a    6682 c    6750 g    6507 t
ORIGIN


alignment_scores:
           Quality: 1385.00              Length:    681
             Ratio:  2.934                Gaps:     12
 Percent Similarity:  69.310   Percent Identity: 42.291

alignment_block:
 US-09-836-410A-1 x AC015346   ..

 Align seg 1/1  to: AC015346  from: 1  to: 26121

      5 LeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValPr 21
        ::::::||||||:::          :::||||||||||:::|||   ||
   9325 GTGTCGATCTATCGCGTGTTCCAGGAGCAGTATCCGCGCGCCCTGTGCCC 9374

     21 oArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysGluCysL 38
        ||||:::|||||||||||||:::   :::|||:::::::|||:::
   9375 TCGCCGCCTGCCTTTGAACATCGCCAACGGCGACGAGTTTCGCGTCGTGA 9424

     38 euAspArgPheLeuArgMetAsnPheSerLysGlyCysProProValPhe 54
          |||:::::::||||||   ::::::   ||||||   |||||||:::|||
   9425 CCGACGAGTACCTGCGCCGCGGTCTGCGTAAGGGCATTCCGCCGCTATTT 9474

     55 AsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIleValGl 71
        :::::::|||:::|||:::::::      |||:::::::|||:::::::||
   9475 GTCAACGTGCGCACTCTGCACCAGATACCGGAGAGGGCGGCCGTTATCGA 9524

     71 uGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPheA 88
        |||||||::::::   |||      :::|||                |||:
   9525 GGAGCTGGCGCTGCAGTATTTCGAGAACCTTACCCGTTCCGGCCACTTTT 9574

     88 snProAsnAspAspGly.........LysGluGluProProThrThrLeu 101
        ::   :::|||                          ||||||   ::::::|||
   9575 CTCGCGAAGATGCCGACGCCGGAATTCCCGTCGAGCCGGCCTCGGCGCTG 9624

    102 LeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnPr 118
        :::|||:::      :::||||||||||||||||||||  :::   :::
   9625 GTGTGGACGGCACTGTTTCTGGCGCAGCACTACGACTACATGCGCGATAC 9674

    118 oSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrL 135
         :::   |||||||||||||||||||:::|||||||:::   ||||||||||
   9675 GGACCGCGCTCTGGAGTACATCAATGTGGCTATCGACCATACGCCAACAC 9724

    135 euIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsn 151
        ||||||||||:::::::::|||:::::|||:::|||||||||||||:::
   9725 TCATTGAGCTGCTTATCACCAAGGGTCGCATCTTTAAGCATGCTGGCGAT 9774

    152 IleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAl 168
              ||||||      |||:::::::||||||||:::::::||||||||
   9775 CCCGTGGAGGCGTACGTCTGGCTGGAGGAGGCCCAAAGCATGGACACGGC 9824

    168 aAspArg..................................... 170
        |||||||
   9825 AGATCGGTGAGTGAGCCCCATGGACTAGCATAGACCTGCTGAACCATGTA 9874

    171 .....................PheIleAsnSerLysCysAlaLysTyr 179
                             :::||||||||||||||||||||||||
   9875 ACCATCAATGTTTTCCCCTCAGCTACATCAATTCGAAGTGCGCGAAGTAC 9924

    180 MetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerLysPh 196
        ||||||:::||||||:::::::::||||||||||||:::|||:::|||||
   9925 ATGCTGCGCGCCAACATGGTGCAGGAGGCAGAGGAGATCTGCGCCAAGTT 9974

    196 eThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGlnCysM 213
        |||||||||||||:::||||||::::::|||||||||||||||||||||||
   9975 CACTCGCGAGGGTGTCTCCGCCATGGACAACTTGAACGAGATGCAGTGCA 10024

    213 etTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnLysPhe 229
        ||||||||||||||||||||||   |||||||:::   |||:::::::::
  10025 TGTGGTTCCAGACGGAGTGTGCCCTGGCCTATCAGCGCATGGGTCGCTGG 10074

    230 GlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIleGluIl 246
        ||||||:::|||||||||||||||||||:::||||||||||||   |||||
  10075 GGCGAGTCGTTGAAGAAGTGCCACGAGGTGGAGCGCCACTTTGCCGAAAT 10124

    246 eThrAspAspGlnPheAspPheHisThrTyrCysMetArgLysIleThrL 263
        |:::::|||||||||||||||||||||||||||||||||||||:::||||
  10125 CGTCGAGGACCAGTTCGATTTCCACACCTATTGCATGCGTAAAATGACGC 10174

    263 euArgSerTyrValAspLeuLeuLysLeuGluAspValLeuArgGlnHis 279
        |||||:::||||||   ||||||:::|||||||||||||||||||||||||
  10175 TGCGCGCCTACGTTGGCCTGTTGCGATTGGAGGACGTGCTACGCCAGCAT 10224

    280 ProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyrLeuLysLe 296
        |||||||||||||||||||:::   ||||||||||:::|||:::::::||
  10225 CCATTCTATTTCAAGGCGGCCAAGTGCGCCATCGAGGTGTACATTCGTCT 10274

    296 uHisAspAsnProLeuThrAspGluAsnLysGluHisGluAlaAsp.... 311
        |:::|||:::||||||   :::|||:::      :::|||   |||
  10275 GTACGACAAGCCGCTTAAGTCGGAGACAACCATTGAAGAGATTGACATTG 10324

    311 ..................................................... 311

  10325 GTATGTGGACATTACCCCACTAGCTCCCTTGCTCGTTCTAGTTAGCTGTA 10374

    312 .................ThrAlaAsnMetSerAspLysGluLeuLysLy 322
                            |||   |||:::      :::||||||||||||
  10375 TGACCGCCTTTCTTCGTTACAGAGAACCTGCCGCCATCAGAACTGAAGAA 10424

    322 sLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGlu. 338
        |||||||:::||||||||||:::|||:::||||||||||:::::|||:::
  10425 GCTGCGCAGCAAACAGCGCAAGGCCAAGAAGAAAGCTGAGCTGGAGAGTG 10474

    339 ..GluLysLysAsnAlaGluLysGluLysProGlnArgAsnProLysLys 354
          :::   :::|||:::  ::::::   |||::::::   :::|||
  10475 CGCAGGCCGCACAGGCGCAGGTGAAGCGCGAGCAGCACCAGAAATCGAAA 10524

    355 LysLysAspAspAspAspGluGluIleGlyGlyProLys...GluGluLe 370
        ::::::   :::::::::   :::      :::|||:::   :::||||||
  10525 CAGCAGGCAAACCAGGAGACCGACCCCGATGCTCCGCAGTTGGACGAGCT 10574

    370 uIleProGluLysLeuAlaLysValGluThrProLeuGluGluAlaIleL 387
        |:::   |||||||||   :::::::::   ||||||:::::||||||:
  10575 GGTGGCCGAGAAGCTGGAGCGCACGGACGATCCGCTGGACAAGGCCATTG 10624

    387 ysPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluThrHis 403
        ::||||||   ||||||:::::::|||:::|||:::::|||||||||||||
  10625 AATTCTTGAAACCGCTGCAGCAGCTGGCTAAGGAGCGCATCGAGACGCAT 10674

    404 LeuPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeuMetLe 420
        |||:::|||||||||:::|||:::|||||||:::|||:::||||||||||||
  10675 CTGCTGGCCTTCGAGTTGTACTACCGCAAGAACAAACTGCTCCTTATGCT 10724

    420 uGlnSerValLysArgAlaPheAlaIleAspSerSerHisProTrpLeuH 437
        ||||   ::::::||||||   |||:::|||:::||||||||   :::|
  10725 GCAGTGCATCCGGCGCGCTCGCGCCGTCGACGCCTCGCACCCCGTAATCC 10774
```

```
   437 isGluCysMetIleArgLeuPheHisSerValCysGlu...SerLysAsp 452
       ||:::|||:::|||||||:::       |||:::   :::   :::|||:::
 10775 ACAGCTGCATTATTCGCTTCGTTAAATCGTTGACCAGCGCCGCCAAGGAG 10824

   453 LeuPro......GluThrValArgThrValLeuLysGlnGluMetAsnAr 467
          |||      |||   |||:::   ||||||::::::      :::::
 10825 CAGCCGTTCAATGAGCACGTGCAGCAGGTGCTGGAGAAGGCCACCAAAGA 10874

   467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
       :|||:::|||:::   ::::|||::::::::::|||:::   |||:::   :
 10875 ACTGATAGGCAGCAAGACGCCCCAGCAGCTCAACGATGAGTTTATTGCCA 10924

   484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
       ::::::::   |||:::   |||      :::::::|||:::   :::|||
 10925 AACACAACGCTTCCATACTGCATTTATACGAGGGCGCACGCAGTCTGTAC 10974

   501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
          ||||||:::|||:::::      ||||||:::|||:::|||:::::
 10975 GAACTGGACAATAGTAAAAAAGCTGCCGCCATCAAGCTGGTCACTAGCTT 11024

   517 uAsp.....................................GlySerLeuT 522
       ::::                                        ::::
 11025 CAACCTGGCCAAGCTCCGGCTGGAGGTAAGTGACTGCCAAACCCATTTCT 11074

   522 hrAsnArgAsnLeuGlnThrCysMetGluValLeuGluAlaLeuCysAsp 538
       ::::::::   |||      |||   ::::::|||      ||||||:::
 11075 CCACCCACTGACTGGCCCCCTGTTGATCCTTGCTCTTTGAATTGTGTTCC 11124

   539 Gly.....................Ser.LeuArgAsp............. 543
       :::                     ||| |||||||||
 11125 GCAGGAGGCCACCAAAATCTACACAGCGTTGCGGGACGGCGATGTGTTTG 11174

   544 .....CysLysGluAlaAlaGluAlaTyrArgAlaSerCysHisLysLeu 558
            |||:::      |||   :::|||:::   :::|||||||:::
 11175 GCGAGTGCGAGGCAGAGGCCGCTTCCTACCAGCAGGCGTGCCACCAACGC 11224

   559 PheProTyrAlaLeuAlaPhe............................ 565
       |||   ||||||      |||
 11225 TTCCAGTACGCCCGCATCTTTCGCAATGTCGAGGAGCTGGAGGCTCAACT 11274

   566 ................MetProProGlyTyrGluGluAspMetLysIleT 577
                                 ||||||:::::::::
 11275 GCAGGAGAAGGAGGCCGCCAAGCTGCGCGCCGAGGAAGAGCAGCAGCAGC 11324

   577 hrValAsnGlyAspSerSerAlaGluThrGluGluLeuAla 590
          ::::::   |||||||||      ::::::   ||||||
 11325 TGATTCACGTTGATTCTAGTGAACCGGTGTCTGTTTTGGCC 11365
```

```
seq_name: gb_in:AC010847

seq_documentation_block:
LOCUS       AC010847              180213 bp    DNA     linear   INV 17-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome X, region 18D-18D, BAC clone
            BACR10M08, complete sequence.
ACCESSION   AC010847
VERSION     AC010847.11  GI:13374650
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 180213)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome X, region 18D-18D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180213)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 17, 2001 this sequence version replaced gi:6563418.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .180213
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="X"
                      /map="18D-18D"
                      /clone="BACR10M08 (D904)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    47472 a  42674 c  42461 g  47606 t
ORIGIN


alignment_scores:
          Quality: 1385.00              Length:    681
            Ratio:  2.934                 Gaps:     12
 Percent Similarity:  69.310   Percent Identity: 42.291

alignment_block:
 US-09-836-410A-1 x AC010847   ..

 Align seg 1/1  to: AC010847  from: 1  to: 180213

       5 LeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValPr 21
         ::::::||||||:::             :::||||||||||:::|||   ||
   23030 GTGTCGATCTATCGCGTGTTCCAGGAGCAGTATCCGCGCGCCCTGTGCCC 23079

      21 oArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysGluCysL 38
         ||||:::|||||||||||||:::   :::|||:::::::|||:::
   23080 TCGCCGCCTGCCTTTGAACATCGCCAACGGCGACGAGTTTCGCGTCGTGA 23129

      38 euAspArgPheLeuArgMetAsnPheSerLysGlyCysProProValPhe 54
           |||:::::|||||   ::::::   ||||||   ||||||:::|||
   23130 CCGACGAGTACCTGCGCCGCGGTCTGCGTAAGGGCATTCCGCCGCTATTT 23179

      55 AsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIleValGl 71
         ::::::|||:::|||:::::       |||:::::::|||:::::||
   23180 GTCAACGTGCGCACTCTGCACCAGATACCGGAGAGGGCGGCCGTTATCGA 23229

      71 uGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPheA 88
         |||||||::::   |||      :::|||               |||:
   23230 GGAGCTGGCGCTGCAGTATTTCGAGAACCTTACCCGTTCCGGCCACTTTT 23279
```



seq_name: gb_in:AC011071

```
seq_documentation_block:
LOCUS       AC011071              182623 bp    DNA     linear   INV 30-MAY-2001
DEFINITION  Drosophila melanogaster, chromosome X, region 18C-18D, BAC clone
            BACR27L16, complete sequence.
ACCESSION   AC011071
VERSION     AC011071.12  GI:14249062
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 182623)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome X, region 18C-18D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182623)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On May 30, 2001 this sequence version replaced gi:7143400.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .182623
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="18C-18D"
                     /clone="BACR27L16 (D903)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    49691 a  41087 c  41478 g  50367 t
ORIGIN


alignment_scores:
            Quality: 1385.00            Length:    681
              Ratio:  2.934              Gaps:     12
 Percent Similarity:  69.310   Percent Identity: 42.291
```

```
alignment_block:
 US-09-836-410A-1 x AC011071   ..

 Align seg 1/1  to: AC011071  from: 1  to: 182623

      5 LeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValPr 21
        :::::|||||||:::           :::||||||||||:::|||   ||
  84550 GTGTCGATCTATCGCGTGTTCCAGGAGCAGTATCCGCGCGCCCTGTGCCC 84599

     21 oArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysGluCysL 38
        ||||:::|||||||||||||:::   :::|||:::::|||:::
  84600 TCGCCGCCTGCCTTTGAACATCGCCAACGGCGACGAGTTTCGCGTCGTGA 84649

     38 euAspArgPheLeuArgMetAsnPheSerLysGlyCysProProValPhe 54
        ||||:::::|||||   ::::::   ||||||   |||||||:::|||
  84650 CCGACGAGTACCTGCGCCGCGGTCTGCGTAAGGGCATTCCGCCGCTATTT 84699

     55 AsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIleValGl 71
        :::::|||:::|||::::::      |||:::::|||:::::||
  84700 GTCAACGTGCGCACTCTGCACCAGATACCGGAGAGGGCGGCCGTTATCGA 84749

     71 uGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPheA 88
        |||||||::::::   |||      :::|||                |||:
  84750 GGAGCTGGCGCTGCAGTATTTCGAGAACCTTACCCGTTCCGGCCACTTTT 84799

     88 snProAsnAspAspGly.....LysGluGluProProThrThrLeu 101
        ::   :::|||                    ||||||   ::::::|||
  84800 CTCGCGAAGATGCCGACGCCGGAATTCCCGTCGAGCCGGCCTCGGCGCTG 84849

    102 LeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnPr 118
        :::|||:::      :::||||||||||||||||||   :::   :::
  84850 GTGTGGACGGCACTGTTTCTGGCGCAGCACTACGACTACATGCGCGATAC 84899

    118 oSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrL 135
        :::   |||||||||||||||||||:::|||||:::   ||||||||||
  84900 GGACCGCGCTCTGGAGTACATCAATGTGGCTATCGACCATACGCCAACAC 84949

    135 euIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsn 151
        ||||||||||:::::::::|||:::::|||:::||||||||||||:::
  84950 TCATTGAGCTGCTTATCACCAAGGGTCGCATCTTTAAGCATGCTGGCGAT 84999

    152 IleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAl 168
            ||||||      |||:::::::||||||||:::::::||||||||
  85000 CCCGTGGAGGCGTACGTCTGGCTGGAGGAGGCCCAAAGCATGGACACGGC 85049

    168 aAspArg..................................... 170
        |||||||
  85050 AGATCGGTGAGTGAGCCCCATGGACTAGCATAGACCTGCTGAACCATGTA 85099

    171 .....................PheIleAsnSerLysCysAlaLysTyr 179
                             :::|||||||||||||||||||||||||
  85100 ACCATCAATGTTTTCCCCTCAGCTACATCAATTCGAAGTGCGCGAAGTAC 85149

    180 MetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerLysPh 196
        ||||||:::||||||:::::::::|||||||||||||:::|||:::||||
  85150 ATGCTGCGCGCCAACATGGTGCAGGAGGCAGAGGAGATCTGCGCCAAGTT 85199

    196 eThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGlnCysM 213
        ||||||||||||:::||||||:::::|||||||||||||||||||||||
  85200 CACTCGCGAGGGTGTCTCCGCCATGGACAACTTGAACGAGATGCAGTGCA 85249

    213 etTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnLysPhe 229
        ||||||||||||||||||||   ||||||:::   |||:::::::::
  85250 TGTGGTTCCAGACGGAGTGTGCCCTGGCCTATCAGCGCATGGGTCGCTGG 85299

    230 GlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIleGluIl 246
        ||||||:::|||||||||||||||||||:::||||||||||||   |||||
  85300 GGCGAGTCGTTGAAGAAGTGCCACGAGGTGGAGCGCCACTTTGCCGAAAT 85349

    246 eThrAspAspGlnPheAspPheHisThrTyrCysMetArgLysIleThrL 263
        |:::::|||||||||||||||||||||||||||||||||||||||:::||||
  85350 CGTCGAGGACCAGTTCGATTTCCACACCTATTGCATGCGTAAAATGACGC 85399
```

```
  263 euArgSerTyrValAspLeuLeuLysLeuGluAspValLeuArgGlnHis 279
      ||||||:::|||||||   ||||||:::|||||||||||||||||||||||
85400 TGCGCGCCTACGTTGGCCTGTTGCGATTGGAGGACGTGCTACGCCAGCAT 85449

  280 ProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyrLeuLysLe 296
      ||||||||||||||||||||:::   ||||||||||:::|||:::::||
85450 CCATTCTATTTCAAGGCGGCCAAGTGCGCCATCGAGGTGTACATTCGTCT 85499

  296 uHisAspAsnProLeuThrAspGluAsnLysGluHisGluAlaAsp.... 311
      |:::|||:::|||||||   :::|||:::      :::|||   |||
85500 GTACGACAAGCCGCTTAAGTCGGAGACAACCATTGAAGAGATTGACATTG 85549

  311 ..................................................... 311

85550 GTATGTGGACATTACCCCACTAGCTCCCTTGCTCGTTCTAGTTAGCTGTA 85599

  312 .................ThrAlaAsnMetSerAspLysGluLeuLysLy 322
                      |||   |||:::      :::|||||||||||
85600 TGACCGCCTTTCTTCGTTACAGAGAACCTGCCGCCATCAGAACTGAAGAA 85649

  322 sLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGlu. 338
      |||||||:::|||||||||:::|||:::|||||||||:::::::|||:::
85650 GCTGCGCAGCAAACAGCGCAAGGCCAAGAAGAAAGCTGAGCTGGAGAGTG 85699

  339 ..GluLysLysAsnAlaGluLysGluLysProGlnArgAsnProLysLys 354
        :::      :::|||:::   ::::::   |||:::::   :::|||
85700 CGCAGGCCGCACAGGCGCAGGTGAAGCGCGAGCAGCACCAGAAATCGAAA 85749

  355 LysLysAspAspAspAspGluGluIleGlyGlyProLys...GluGluLe 370
      ::::::   :::::::::   :::      :::|||:::   :::|||||
85750 CAGCAGGCAAACCAGGAGACCGACCCCGATGCTCCGCAGTTGGACGAGCT 85799

  370 uIleProGluLysLeuAlaLysValGluThrProLeuGluGluAlaIleL 387
      |:::   |||||||||   :::::::::   ||||||::::::||||||:
85800 GGTGGCCGAGAAGCTGGAGCGCACGGACGATCCGCTGGACAAGGCCATTG 85849

  387 ysPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluThrHis 403
      ::||||||   ||||||:::::::|||:::|||:::::::||||||||||
85850 AATTCTTGAAACCGCTGCAGCAGCTGGCTAAGGAGCGCATCGAGACGCAT 85899

  404 LeuPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeuMetLe 420
      |||:::||||||||||:::|||:::||||||:::|||:::||||||||||
85900 CTGCTGGCCTTCGAGTTGTACTACCGCAAGAACAAACTGCTCCTTATGCT 85949

  420 uGlnSerValLysArgAlaPheAlaIleAspSerSerHisProTrpLeuH 437
      ||||   ::::::||||||   |||:::|||:::||||||||||   :::|
85950 GCAGTGCATCCGGCGCGCTCGCGCCGTCGACGCCTCGCACCCCGTAATCC 85999

  437 isGluCysMetIleArgLeuPheHisSerValCysGlu...SerLysAsp 452
      ||:::|||:::|||||||:::      |||:::   :::   :::|||:::
86000 ACAGCTGCATTATTCGCTTCGTTAAATCGTTGACCAGCGCCGCCAAGGAG 86049

  453 LeuPro......GluThrValArgThrValLeuLysGlnGluMetAsnAr 467
        |||      |||   |||:::   ||||||:::::      :::::
86050 CAGCCGTTCAATGAGCACGTGCAGCAGGTGCTGGAGAAGGCCACCAAAGA 86099

  467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
      :|||:::|||:::   :::|||:::::::::|||:::   |||:::  :
86100 ACTGATAGGCAGCAAGACGCCCCAGCAGCTCAACGATGAGTTTATTGCCA 86149

  484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
      ::::::::   |||:::   |||      ::::::|||:::   :::|||
86150 AACACAACGCTTCCATACTGCATTTATACGAGGGCGCACGCAGTCTGTAC 86199

  501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
      |||||||:::|||::::::      ||||||:::|||:::|||:::::
86200 GAACTGGACAATAGTAAAAAAGCTGCCGCCATCAAGCTGGTCACTAGCTT 86249

  517 uAsp.....................................GlySerLeuT 522
      ::::                                        ::::
86250 CAACCTGGCCAAGCTCCGGCTGGAGGTAAGTGACTGCCAAACCCATTTCT 86299

  522 hrAsnArgAsnLeuGlnThrCysMetGluValLeuGluAlaLeuCysAsp 538
```

```
      :::::::   |||      |||   ::::::|||      ||||||:::
86300 CCACCCACTGACTGGCCCCCTGTTGATCCTTGCTCTTTGAATTGTGTTCC 86349

  539 Gly.....................Ser.LeuArgAsp............. 543
      :::                     ||| |||||||||
86350 GCAGGAGGCCACCAAAATCTACACAGCGTTGCGGGACGGCGATGTGTTTG 86399

  544 .....CysLysGluAlaAlaGluAlaTyrArgAlaSerCysHisLysLeu 558
           |||:::      |||   :::|||:::   :::|||||||:::
86400 GCGAGTGCGAGGCAGAGGCCGCTTCCTACCAGCAGGCGTGCCACCAACGC 86449

  559 PheProTyrAlaLeuAlaPhe........................... 565
      |||   ||||||      |||
86450 TTCCAGTACGCCCGCATCTTTCGCAATGTCGAGGAGCTGGAGGCTCAACT 86499

  566 ..............MetProProGlyTyrGluGluAspMetLysIleT 577
                                    ||||||:::::::::
86500 GCAGGAGAAGGAGGCCGCCAAGCTGCGCGCCGAGGAAGAGCAGCAGCAGC 86549

  577 hrValAsnGlyAspSerSerAlaGluThrGluGluLeuAla 590
        ::::::   |||||||||      ::::::   ||||||
86550 TGATTCACGTTGATTCTAGTGAACCGGTGTCTGTTTTGGCC 86590
```

```
seq_name: gb_in:AE003512

seq_documentation_block:
LOCUS       AE003512              301457 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386053 section 29
            of 30, complete sequence.
ACCESSION   AE003512 AE002593
VERSION     AE003512.2  GI:10728339
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 301457)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
```

```
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 301457)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7293579.
FEATURES             Location/Qualifiers
     source          1. .301457
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(<54416. .54439,54540. .>54818)
                     /gene="CG14197"
                     /product="CT33810"
                     /db_xref="FLYBASE:FBan0014197"
                     /db_xref="FLYBASE:FBgn0031015"
                     /evidence=not_experimental
     gene            <54416. .>54818
                     /gene="CG14197"
                     /map="18C8-18C8"
                     /db_xref="FLYBASE:FBan0014197"
                     /db_xref="FLYBASE:FBgn0031015"
                     /evidence=not_experimental
     CDS             join(54416. .54439,54540. .54818)
                     /gene="CG14197"
                     /note="CG14197 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0014197"
                     /db_xref="FLYBASE:FBgn0031015"
                     /evidence=not_experimental
                     /protein_id="AAF48952.1"
                     /db_xref="GI:7293580"
                     /translation="MGEMDGGGDENEDGDEDEDEDEDEDEDEDEDEDEDEDEDEDE
                     DEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDGTGTGTGDS"
     mRNA            join(66369. .67950,70036. .71770)
                     /gene="CG12199"
                     /product="CT10486"
                     /db_xref="FLYBASE:FBan0012199"
                     /db_xref="FLYBASE:FBgn0031016"
     gene            <66369. .>71770
                     /gene="CG12199"
                     /map="18C8-18C8"
                     /db_xref="FLYBASE:FBan0012199"
                     /db_xref="FLYBASE:FBgn0031016"
     CDS             join(66516. .67950,70036. .71279)
                     /gene="CG12199"
                     /note="CG12199 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012199"
                     /db_xref="FLYBASE:FBgn0031016"
                     /protein_id="AAF48953.1"
                     /db_xref="GI:7293581"
                     /translation="MQSCGTCHCQWNSGKKSADCKNKALTKIPQDMSNEMQVLDFAHN
                     QIPELRREEFLLAGLPNVHKIFLRNCTIQEVHREAFKGLHILIELDLSGNRIRELHPG
                     TFAGLEKLRNVIINNNEIEVLPNHLFVNLSFLSRIEFRNNRLRQVQLHVFAGTMALSA
                     ISLEQNRLSHLHKETFKDLQKLMHLSLQGNAWNCSCELQDFRDFAISKRLYTPPTDCQ
                     EPPQLRGKLWSEVPSENFACRPRILGSVRSFIEANHDNISLPCRIVGSPRPNVTWVYN
                     KRPLQQYDPRVRVLTSVEQMPEQPSQVLTSELRIVGVRASDKGAYTCVADNRGGRAEA
                     EFQLLVSGDYAGAVSASDGMGMGAIGAPTIDPQTNMFLIICLIITTLLLLLLVAVLTL
                     FWYCRRIKTYQKDTTMMSGDGLISSKMDKTHNGSMLEGSVIMEMQKSLLNEVNPVEKP
                     PRRTDIESVDGGDDVLEIKKTLLDDTVYVANHSRDEEAVSVAMSDTTTTPRSRHTYVD
                     DAYANSLPPDLLAFPARVPPTSPSMQSSQSNIPDQVIYGIRSPPSLTSPVYTHMTPHG
                     IYGTKTMTAPHNGFMTLQHPKSRNLALIATTNSSRQHQHHHQLQQQQQHHHHHQQQQQ
                     QQQQQQHPLATTSPFLPAPVVYSPATGVVMKQGYMTIPRKPRAPSWAPSTSGAAGHGS
                     IQLSEFQSPTSPNPSETGTATTAELQAEPVYDNLGLRTTAGGNSTLNLTKIAGSQGGA
                     GQQYSMRDRPLPATPSLTSVSSATNASKIYEPIHELIQQQQQLQQQQQQQQQRLGSMD
                     TEPLYGVRQQGITILPGSSISGAGLGHAAYLSPGSGAAVSPSHASSSGDSPKAAKIPP
                     RPPPKPKKKMSVTTTRSGQGSTSQLFDDEGEDGTEV"
     mRNA            complement(<75911. .>76414)
                     /gene="CG14198"
                     /product="CT33811"
                     /db_xref="FLYBASE:FBan0014198"
                     /db_xref="FLYBASE:FBgn0031017"
                     /evidence=not_experimental
     gene            complement(<75911. .>76414)
                     /gene="CG14198"
                     /map="18C8-18D1"
                     /db_xref="FLYBASE:FBan0014198"
                     /db_xref="FLYBASE:FBgn0031017"
                     /evidence=not_experimental
     CDS             complement(75911. .76414)
                     /gene="CG14198"
                     /note="CG14198 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0014198"
                     /db_xref="FLYBASE:FBgn0031017"
                     /evidence=not_experimental
                     /protein_id="AAF48954.1"
                     /db_xref="GI:7293582"
                     /translation="MALDVAYAIRRLLPADIKRLYTHQAPSARLAELKENPFAVDFPM
                     TPNEEKGGLNVSEHVVYSCLAEQQWTTAMDATLRAEPWQCTRCDFELEEFDVLEQLVH
                     RSKCKRRKAQESRTAKTTSTESSTEVASSSSSHSSGGYYCNSCKRELRLTTIDILRHK
                     KQCRNNK"
     mRNA            76856. .77551
                     /gene="CG12200"
                     /product="CT10592"
                     /db_xref="FLYBASE:FBan0012200"
                     /db_xref="FLYBASE:FBgn0031018"
     gene            <76856. .>77551
                     /gene="CG12200"
                     /map="18C8-18D1"
                     /db_xref="FLYBASE:FBan0012200"
                     /db_xref="FLYBASE:FBgn0031018"
     CDS             76896. .77531
                     /gene="CG12200"
                     /note="CG12200 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012200"
                     /db_xref="FLYBASE:FBgn0031018"
                     /protein_id="AAF48955.1"
                     /db_xref="GI:7293583"
                     /translation="MFRVHCNKCFRHRKTDPAVPFHLTQCRHVICGPCLGQSSLEKNC
                     PLCGQVLKAIQINRDMPTSVANYFADPLRFQQIYRKISKFQADQRASDNLGFYRQLQQ
                     LEQNKRQLEGFCKMEAQLNQKVVEEKKRIAELRTYIAYHENAQRMTRRRHSAGERFHT
                     PEFKEAWNTSISTSDKSPSDMPSDSSRRSADLDTQSTRRRSFGSDTKGFRL"
     mRNA            complement(join(77679. .78071,78140. .79068,79137. .79334,
                     79397. .79767,79822. .80352,80413. .81049,81201. .81460))
                     /gene="CG12211"
                     /product="CT11197"
                     /db_xref="FLYBASE:FBan0012211"
                     /db_xref="FLYBASE:FBgn0031019"
     gene            complement(<77679. .>81460)
                     /gene="CG12211"
                     /map="18C8-18D1"
                     /db_xref="FLYBASE:FBan0012211"
                     /db_xref="FLYBASE:FBgn0031019"
     CDS             complement(join(77771. .78071,78140. .79068,79137. .79334,
                     79397. .79767,79822. .80352,80413. .81049,81201. .81446))
                     /gene="CG12211"
                     /note="CG12211 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012211"
                     /db_xref="FLYBASE:FBgn0031019"
                     /protein_id="AAF48956.1"
                     /db_xref="GI:7293584"
                     /translation="MIRDVKRRKCVKENAVERLRVNSLRKNDILPPELREVADAEIAA
                     FPRDSSLVRVRERCALTSRPRGVVHKYRLSRIVWRHLACYRLLWEEVNTNFRNYFASV
                     KMSRIKKKADEHASSRRRTQINLIEFSFLDHKAELVGLLQAKSRQVGRSLVENEADFW
                     KFVGKYEAMMRSTGQPVLPPRLAEAELSSMQPYHRSKHLALRLDADELLQEARDQMQS
                     QFRHILLVYLDFRQTEKFQRIRKLRQTQRNLPIARFRKDLREALDTSRVVIVAGDTGC
```

```
                    GKSTQVPQYLYDFGYRSIACTQPRRLACVSLCKRVAHELLDDYGSRVAFQIRFERSRT
                    KLTNILFITEGLLLRQLAVAANLDQYDALILDEIHERNLFGDFLLGVTKCLLRARPQL
                    KLILMSATINVELFHGYFGEEGARLVQVPGRLFPIKLRYLPPPALELKAGQATSKRSQ
                    RNRIDPAPFVQVLSLIDQQYPTSERGDVLIFVSGVNEIESVVEAVHEYATEQTHWLVL
                    PLHSGQAIADQSKVFDYAPEGMRKCIVSTNIAETSLTVDGVRFVVDSGKVKEMNFDAT
                    CKGQRLKEFWVSKSSADQRKGRAGRTGPGVCFRLYTAEQYNAFEAYPTPEIYRVPLDT
                    MLLQMVSMGLPDVRAFPFIEAPETERIEQTILALKQHCALSVEEKITPLGRSLANLPV

alignment_scores:
           Quality: 1385.00              Length:    681
             Ratio:  2.934                Gaps:     12
 Percent Similarity:  69.310    Percent Identity: 42.291

alignment_block:
 US-09-836-410A-1 x AE003512   ..

 Align seg 1/1  to: AE003512  from: 1  to: 301457

      5 LeuLysIleTyrGluGluAlaTrpThrLysTyrProArgGlyLeuValPr 21
        ::::::|||||:::             :::|||||||||:::|||   ||
  83609 GTGTCGATCTATCGCGTGTTCCAGGAGCAGTATCCGCGCGCCCTGTGCCC 83658

     21 oArgLysLeuProLeuAsnPheLeuSerGlyGluLysPheLysGluCysL 38
        ||||:::|||||||||||||:::   :::|||::::::|||:::
  83659 TCGCCGCCTGCCTTTGAACATCGCCAACGGCGACGAGTTTCGCGTCGTGA 83708

     38 euAspArgPheLeuArgMetAsnPheSerLysGlyCysProProValPhe 54
          |||:::::::||||||   ::::::   ||||||   ||||||:::|||
  83709 CCGACGAGTACCTGCGCCGCGGTCTGCGTAAGGGCATTCCGCCGCTATTT 83758

     55 AsnThrLeuArgSerLeuTyrArgAspLysGluLysValAlaIleValGl 71
        :::::::|||:::|||::::::      |||:::::|||:::::::||
  83759 GTCAACGTGCGCACTCTGCACCAGATACCGGAGAGGGCGGCCGTTATCGA 83808

     71 uGluLeuValValGlyTyrGluThrSerLeuLysSerCysArgLeuPheA 88
        |||||||::::::   |||      :::|||                |||:
  83809 GGAGCTGGCGCTGCAGTATTTCGAGAACCTTACCCGTTCCGGCCACTTTT 83858

     88 snProAsnAspAspGly.........LysGluGluProProThrThrLeu 101
        ::   :::|||                      ||||||   :::::::|||
  83859 CTCGCGAAGATGCCGACGCCGGAATTCCCGTCGAGCCGGCCTCGGCGCTG 83908

    102 LeuTrpValGlnTyrTyrLeuAlaGlnHisTyrAspLysIleGlyGlnPr 118
        :::|||:::      :::||||||||||||||||||   :::   :::
  83909 GTGTGGACGGCACTGTTTCTGGCGCAGCACTACGACTACATGCGCGATAC 83958

    118 oSerIleAlaLeuGluTyrIleAsnThrAlaIleGluSerThrProThrL 135
        :::   |||||||||||||||||:::|||||:::   |||||||||
  83959 GGACCGCGCTCTGGAGTACATCAATGTGGCTATCGACCATACGCCAACAC 84008

    135 euIleGluLeuPheLeuValLysAlaLysIleTyrLysHisAlaGlyAsn 151
        ||||||||||:::::::|||:::::|||:::||||||||||||:::
  84009 TCATTGAGCTGCTTATCACCAAGGGTCGCATCTTTAAGCATGCTGGCGAT 84058

    152 IleLysGluAlaAlaArgTrpMetAspGluAlaGlnAlaLeuAspThrAl 168
            ||||||      |||:::::|||||||||:::::|||||||||
  84059 CCCGTGGAGGCGTACGTCTGGCTGGAGGAGGCCCAAAGCATGGACACGGC 84108

    168 aAspArg..................................... 170
        |||||||
  84109 AGATCGGTGAGTGAGCCCCATGGACTAGCATAGACCTGCTGAACCATGTA 84158

    171 ......................PheIleAsnSerLysCysAlaLysTyr 179
                               :::|||||||||||||||||||||||||
  84159 ACCATCAATGTTTTCCCCTCAGCTACATCAATTCGAAGTGCGCGAAGTAC 84208

    180 MetLeuLysAlaAsnLeuIleLysGluAlaGluGluMetCysSerLysPh 196
        ||||||:::|||||:::::::::|||||||||||||:::|||:::||||||
  84209 ATGCTGCGCGCCAACATGGTGCAGGAGGCAGAGGAGATCTGCGCCAAGTT 84258

    196 eThrArgGluGlyThrSerAlaValGluAsnLeuAsnGluMetGlnCysM 213
        ||||||||||||||:::||||||:::::|||||||||||||||||||||||
  84259 CACTCGCGAGGGTGTCTCCGCCATGGACAACTTGAACGAGATGCAGTGCA 84308

    213 etTrpPheGlnThrGluCysAlaGlnAlaTyrLysAlaMetAsnLysPhe 229
        ||||||||||||||||||||||   ||||||:::   |||:::::::::
  84309 TGTGGTTCCAGACGGAGTGTGCCCTGGCCTATCAGCGCATGGGTCGCTGG 84358

    230 GlyGluAlaLeuLysLysCysHisGluIleGluArgHisPheIleGluIl 246
        ||||||:::|||||||||||||||||||:::||||||||||   |||||
  84359 GGCGAGTCGTTGAAGAAGTGCCACGAGGTGGAGCGCCACTTTGCCGAAAT 84408

    246 eThrAspAspGlnPheAspPheHisThrTyrCysMetArgLysIleThrL 263
        |:::::|||||||||||||||||||||||||||||||||||||:::||||
  84409 CGTCGAGGACCAGTTCGATTTCCACACCTATTGCATGCGTAAAATGACGC 84458

    263 euArgSerTyrValAspLeuLeuLysLeuGluAspValLeuArgGlnHis 279
        |||||:::||||||   ||||||:::||||||||||||||||||||||||
  84459 TGCGCGCCTACGTTGGCCTGTTGCGATTGGAGGACGTGCTACGCCAGCAT 84508

    280 ProPheTyrPheLysAlaAlaArgIleAlaIleGluIleTyrLeuLysLe 296
        ||||||||||||||||||||:::   |||||||||:::|||:::::||
  84509 CCATTCTATTTCAAGGCGGCCAAGTGCGCCATCGAGGTGTACATTCGTCT 84558

    296 uHisAspAsnProLeuThrAspGluAsnLysGluHisGluAlaAsp.... 311
        |:::|||:::||||||   :::|||:::      :::|||   |||
  84559 GTACGACAAGCCGCTTAAGTCGGAGACAACCATTGAAGAGATTGACATTG 84608

    311 ..................................................... 311

  84609 GTATGTGGACATTACCCCACTAGCTCCCTTGCTCGTTCTAGTTAGCTGTA 84658

    312 .................ThrAlaAsnMetSerAspLysGluLeuLysLy 322
                          |||   |||:::     :::||||||||||||
  84659 TGACCGCCTTTCTTCGTTACAGAGAACCTGCCGCCATCAGAACTGAAGAA 84708

    322 sLeuArgAsnLysGlnArgArgAlaGlnLysLysAlaGlnIleGluGlu. 338
        |||||||:::||||||||||:::|||:::|||||||||:::::|||:::
  84709 GCTGCGCAGCAAACAGCGCAAGGCCAAGAAGAAAGCTGAGCTGGAGAGTG 84758

    339 ..GluLysLysAsnAlaGluLysGluLysProGlnArgAsnProLysLys 354
          :::      :::|||:::   :::::   |||::::::   :::|||
  84759 CGCAGGCCGCACAGGCGCAGGTGAAGCGCGAGCAGCACCAGAAATCGAAA 84808

    355 LysLysAspAspAspAspGluGluIleGlyGlyProLys...GluGluLe 370
        ::::::   :::::::::   :::      :::|||:::   :::||||||
  84809 CAGCAGGCAAACCAGGAGACCGACCCCGATGCTCCGCAGTTGGACGAGCT 84858

    370 uIleProGluLysLeuAlaLysValGluThrProLeuGluGluAlaIleL 387
        |:::  |||||||||   :::::::::   ||||||::::::|||||||:
  84859 GGTGGCCGAGAAGCTGGAGCGCACGGACGATCCGCTGGACAAGGCCATTG 84908

    387 ysPheLeuThrProLeuLysAsnLeuValLysAsnLysIleGluThrHis 403
        ::||||||  ||||||:::::|||:::|||:::::|||||||||||||
  84909 AATTCTTGAAACCGCTGCAGCAGCTGGCTAAGGAGCGCATCGAGACGCAT 84958

    404 LeuPheAlaPheGluIleTyrPheArgLysGluLysPheLeuLeuMetLe 420
        |||:::|||||||||:::|||:::||||||:::|||:::||||||||||||
  84959 CTGCTGGCCTTCGAGTTGTACTACCGCAAGAACAAACTGCTCCTTATGCT 85008

    420 uGlnSerValLysArgAlaPheAlaIleAspSerSerHisProTrpLeuH 437
        ||||  :::::::||||||   |||:::|||:::|||||||||  :::|
  85009 GCAGTGCATCCGGCGCGCTCGCGCCGTCGACGCCTCGCACCCCGTAATCC 85058

    437 isGluCysMetIleArgLeuPheHisSerValCysGlu...SerLysAsp 452
        ||:::|||:::|||||||:::      |||:::  :::  :::|||:::
  85059 ACAGCTGCATTATTCGCTTCGTTAAATCGTTGACCAGCGCCGCCAAGGAG 85108

    453 LeuPro......GluThrValArgThrValLeuLysGlnGluMetAsnAr 467
          |||      |||   |||:::   ||||||::::::      :::::
  85109 CAGCCGTTCAATGAGCACGTGCAGCAGGTGCTGGAGAAGGCCACCAAAGA 85158

    467 gLeuPheGlyAlaThrAsnProLysAsnPheAsnGluThrPheLeuLysA 484
        :|||:::|||:::   :::|||:::::::::|||:::   |||:::  :
  85159 ACTGATAGGCAGCAAGACGCCCCAGCAGCTCAACGATGAGTTTATTGCCA 85208
```

```
  484 rgAsnSerAspSerLeuProHisArgLeuSerAlaAlaLysMetValTyr 500
      ::::::::   |||:::   |||      ::::::|||:::   :::|||
85209 AACACAACGCTTCCATACTGCATTTATACGAGGGCGCACGCAGTCTGTAC 85258

  501 TyrLeuAspSerSerSerGlnLysArgAlaIleGluLeuAlaThrThrLe 517
        ||||||:::|||:::::      ||||||:::|||:::|||:::::
85259 GAACTGGACAATAGTAAAAAAGCTGCCGCCATCAAGCTGGTCACTAGCTT 85308

  517 uAsp.................................GlySerLeuT 522
      ::::                                    ::::
85309 CAACCTGGCCAAGCTCCGGCTGGAGGTAAGTGACTGCCAAACCCATTTCT 85358

  522 hrAsnArgAsnLeuGlnThrCysMetGluValLeuGluAlaLeuCysAsp 538
      ::::::::   |||      |||   ::::::|||      ||||||:::
85359 CCACCCACTGACTGGCCCCCTGTTGATCCTTGCTCTTTGAATTGTGTTCC 85408

  539 Gly.....................Ser.LeuArgAsp............. 543
      :::                      ||| |||||||||
85409 GCAGGAGGCCACCAAAATCTACACAGCGTTGCGGGACGGCGATGTGTTTG 85458

  544 .....CysLysGluAlaAlaGluAlaTyrArgAlaSerCysHisLysLeu 558
           |||:::      |||   :::|||:::   :::|||||||:::
85459 GCGAGTGCGAGGCAGAGGCCGCTTCCTACCAGCAGGCGTGCCACCAACGC 85508

  559 PheProTyrAlaLeuAlaPhe............................ 565
      |||   ||||||      |||
85509 TTCCAGTACGCCCGCATCTTTCGCAATGTCGAGGAGCTGGAGGCTCAACT 85558

  566 ...............MetProProGlyTyrGluGluAspMetLysIleT 577
                              ||||||:::::::::
85559 GCAGGAGAAGGAGGCCGCCAAGCTGCGCGCCGAGGAAGAGCAGCAGCAGC 85608

  577 hrValAsnGlyAspSerSerAlaGluThrGluGluLeuAla 590
        ::::::   |||||||||      ::::::   ||||||
85609 TGATTCACGTTGATTCTAGTGAACCGGTGTCTGTTTTGGCC 85649
```